Sequence Specific Trapping of Topoisomerase I by Camptothecin Polyamide Conjugates

Thesis by

Clay Chia Chun Wang

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(Submitted July 30, 2001)

To my family

Acknowledgements

First and most importantly, I would like to thank my advisor, Professor Peter B. Dervan, for his support over the past few years. His enthusiasm for chemistry is truly inspirational. I am very fortunate to be part of the Dervan research family. I would also like to thank my committee members Professors Doug Rees, Rich Roberts, and Brian Stoltz for their encouragement and advice.

I would not be here without the help and encouragement from my family. I would like to thank my family for helping me fulfill my lifelong dream of being a chemist. Particularly I would like to thank my girlfriend, Tammy, for taking care of me over the long journey. This work is truly a team effort between the both of us.

There are many folks from the Dervan group to thank. First, I would like to thank Ulf Ellervik for whom I have the privilege to collaborate with over the years. His fearlessness and knowledge in chemistry is the reason our collaboration was so fruitful. I have benefited tremendously working next to Aileen Chang. I have never met anyone as nice as Aileen. I need to thank her for her help ranging from letting me use her glassware and chemicals to teaching me how to make polyamides. I would like to thank Eldon Baird, Ryan Bremer, and Nick Wurtz for taking a leadership role in the group during my stay. I don't think people appreciate the amount of extra work they have to do in order for the group to run smoothly. I would like to thank the post docs in Church, Anna Mapp, Paul Floreancig, Tom Minehan, and Philipp Weyerman, for being such great role models. I would like to thank Doan Nguyen for proofreading this thesis. I enjoyed many great lunch conversations with Doan and Jason Belitsky. It has been fun working with Victor Rucker, Shane Foister, Amanda Cashin, Eric Fechter, Adam Kirkstien, and Mike Marques.

Abstract

Polyamides containing *N*-methylpyrrole (Py), *N*-methylimidazole (Im) and 3-hydroxy-1-methylpyrrole (Hp) are cell permeable small molecules that bind as antiparallel pairs in the minor groove of DNA according to a set of "pairing rules" with affinities and specificities for predetermined sequences comparable to DNA-binding proteins. Several of these ring pairings have only been demonstrated when placed internal in a hairpin and not for the terminal position in hairpin polyamides. Several series of eight ring hairpin polyamides with modification at the N-terminal position were synthesized and characterized. We observed that at the terminal position (i) the Py/Py pair functions similar to when placed internally preferring A•T and T•A base pairs over G•C and C•G, (ii) the Hp/Py pair could not distinguish between T•A over A•T possibly due to rotation of the terminal Hp to form an intramolecular hydrogen bond between the 3-hydroxyl hydrogen and the carboxamide oxygen which would orient the key hydroxyl recognition element away from the minor groove. A new aromatic pair, 2-hydroxy-6-methoxybenzamide/1-methylpyrrole was designed and shown to distinguish T•A from A•T base pairs and both from G•C/C•G and (iii) the Py/Im pair in the classic eight ring hairpin motif showed no preference for C•G base pair possibly due to the mispositioning of the Im residue located at the C-terminal end of the four ring polyamide subunit. Targeting of C•G was accomplished by replacing a pyrrole with a flexible β-alanine and setting the imidazole back in register.

Pyrrole-imidazole polyamides that target DNA sequences in the promoter have been shown to inhibit transcription of specific genes in cell culture. When bound to coding region of genes, polyamides do not appear to inhibit gene expression. A possible solution is to design molecule capable of modifying DNA when bound to the coding

region. A series of polyamide-camptothecin conjugates were designed to trap the enzyme Topoisomerase I and induce cleavage at predetermined DNA sites. Cleavage yields were shown to be dependent on linker length between the DNA binding polyamide and the Topo I trapping camptothecin unit with the camptothecin unit with the longest linker showing greatest cleavage yield of over 90%.

Table of Contents

List of Figures

Chapter 1 page

Chapter 2 page

Chapter 3 page

Chapter 4 page

Chapter 5 page

Chapter 6 page

Chapter 7 page

Chapter 8 page

Chapter 1

Introduction

Introduction

Double helical DNA

Deoxyribonucleic acid (DNA) is the medium found in Nature to carry the genetic blueprint of life. DNA is composed of two antiparallel oligodeoxyribonucleic strands associated by hydrogen bonds between the four bases adenine (A), thymine (T), cytosine (C), and guanine (G).[1] The common B-form DNA is characterized by a wide and shallow major groove and a narrow and deep minor groove (Figure 1.1).[2] Sequences are distinguishable by the pattern of hydrogen bond donors and acceptors displayed on the edges of the base pairs and this hydrogen bond display dictates the pairing of the bases such that A pairs with T and G pairs with C (Figure 1.2).

The successful completion of the human genome project has revealed that our genetic information, stored as DNA polymers, consists of three billion bases that code for 25,000 to 30,000 genes.[3] Specific classes of proteins such as repressors and activators interact with DNA to control gene expression (Figure 1.3). Misregulation of gene expression is responsible for numerous cancerous and viral disease states. Small synthetic molecules that bind to specific DNA sequences could be useful as artificial transcription factors to regulate gene expression.

DNA Recognition.

Structural studies of naturally occurring protein DNA complexes have provided insights into designing novel DNA binding molecules. Protein:DNA interactions are characterized by hydrogen bonds, van der Waals interaction and electrostatic interactions. Nature has devised a combinatorial approach for DNA recognition utilizing a twenty amino acid code to form complex tertiary folded structures. Examples of DNA binding protein motifs include, zinc fingers, leucine zippers, and the helix turn helix motifs

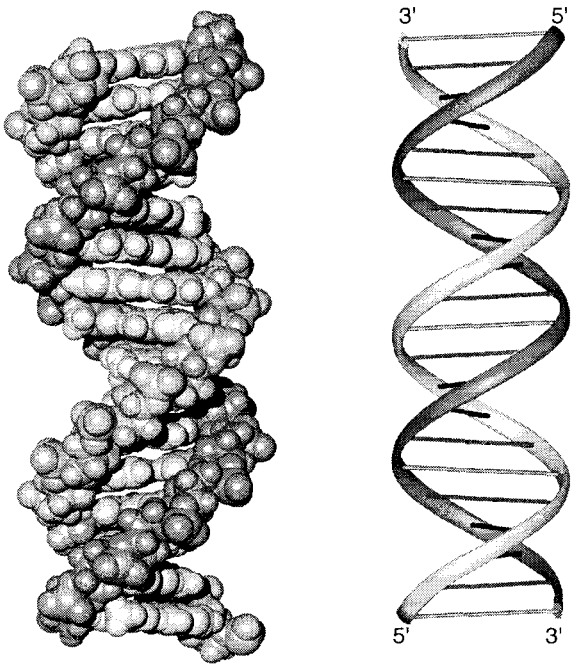


Figure 1.1. B-form double helix DNA. Antiparallel strand are in dark and light grey. (left) space filling CPK model, (right) ribbon representation.

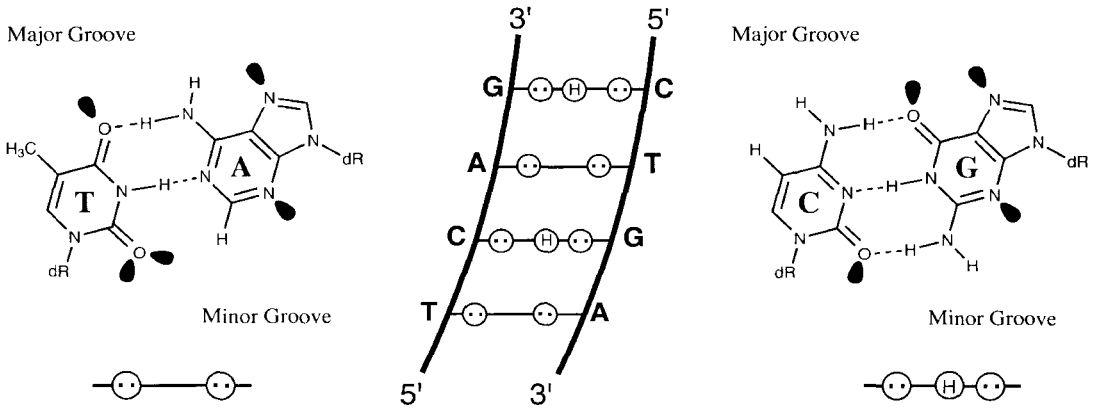


Figure 1.2. A schematic model of recognition of the minor groove, with hydrogen bond alone representative on (H) and hydrogen bond acceptors as two dots.

(Figure 1.4).[4] Within these complexes, specificity for target sites is achieved by noncovalent interactions between the protein side chains, and the nucleobases and phosphates of the DNA. Due to the complexity of protein-DNA interactions, the design of proteins for the recognition of predetermined target has been challenging. A promising approach is the use of phage display to select zinc fingers with desired DNA binding specificities.[5]

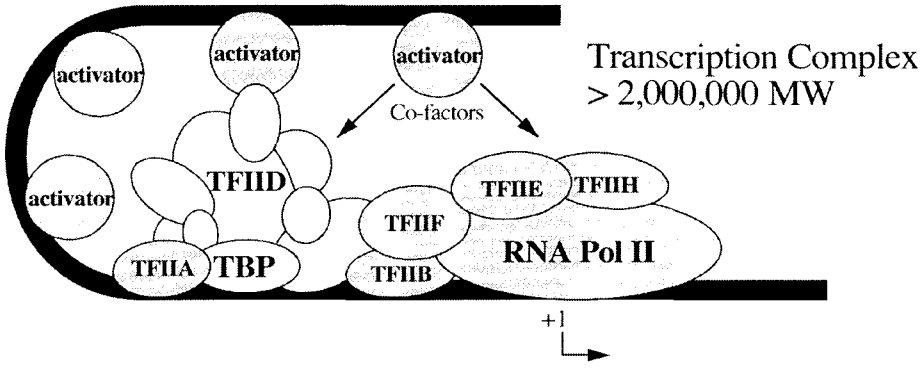


Figure 1.3. Model of protein regulation of gene expression. Initiation of gene expression requires a combinatorial assembly of transcription factors.

Besides proteins, small molecules isolated from Nature form another class of DNA binding ligands. Some examples include chromomycin, distamycin, actinomycin D, and calicheamicin oligosaccharide (Figure 1.5).[6] These molecules bind either in the minor groove or intercalate through the bases. The structural simplicity of distamycin made it an attractive starting point for the design of sequence specific DNA binding small molecule. Distamycin consists of three *N*-methyl carboxamide units and netropsin consists of two *N*-methyl carboxamide units. The crescent shaped molecules bind in the minor groove of DNA at A-T rich regions.

An early goal in the Dervan group was to expand the sequence targeted by distamycin in order to bind G•C base pairs.[7] Various heterocycles were attached to the N-terminal position of distamycin to recognize the sequence 5'-WGWWW-

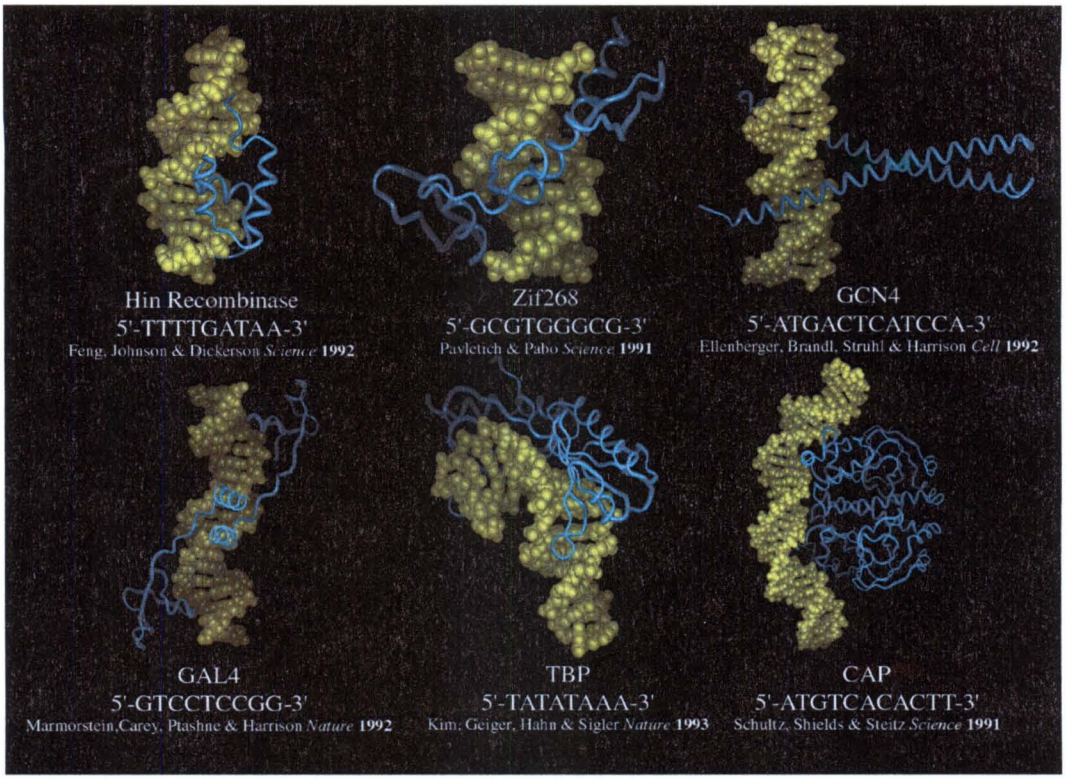
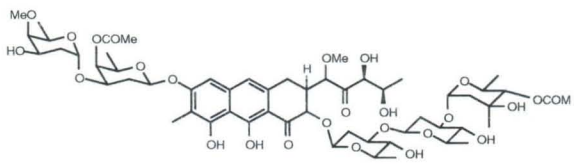
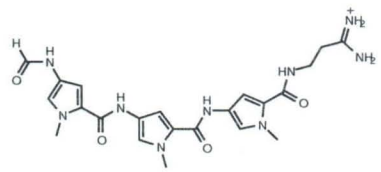


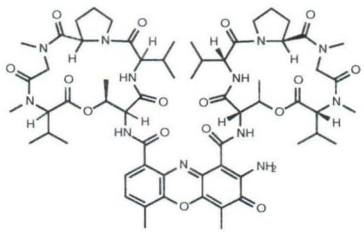
Figure 1.4. X-ray crystal structures from several families of DNA binding proteins. Proteins can contact DNA via interaction with the bases in the major and minor grooves as wells as the sugar phosphate backbone.



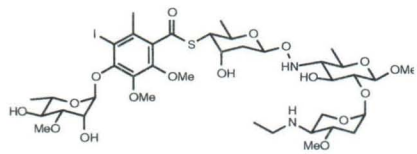
Chromomycin • 5'-GGCT-3'



Distamycin • 5'-$(A,T)_5$-3'



Actinomycin D • 5'-GC-3'



Calicheamicin oligosaccharide • 5'-TCCT-3'

Figure 1.5. Chemical structures of natural products that bind DNA.

3' where W = A or T. Footprinting experiments unexpectedly showed 2-imidazole netropsin bound to the sequence 5'-WGWCW-3'. The twofold symmetry of the binding site was puzzling until NMR structural studies in collaboration with the Wemmer Group showed that at high concentration, distamycin could bind as a 2:1 complex rather than a 1:1 complex (Figure 1.6).[8] The binding of 2-imidazole netropsin to the site 5'-WGWCW-3' as an antiparallel dimer allowed the development of a set of pairing rules for the recognition by *N*-methylpyrrole (Py), *N*-methylimidazole (Im) and *N*-methyl-3-hydroxypyrrole (Hp) amino acids.[9] A Py/Im pair targets a C•G base pair while an Im/Py pair targets a G•C base pair. The basis for discrimination of a G•C base pair is from the formation of a hydrogen bond between imidazole N3 and the exocyclic amine of guanine. A Py/Py pair is partially degenerate and recognizes both A•T and T•A base pairs.[10] High resolution NMR and crystal structures have confirmed that specific hydrogen bonds are made from the carboxamide nitrogens to the base pairs and the imidazole nitrogen to guanine. The use of Hp/Py and Py/Hp pairs for discrimination of T•A and A•T base pairs, respectively, completes the code.[11] Specificity for A, T base pairs is accomplished by a

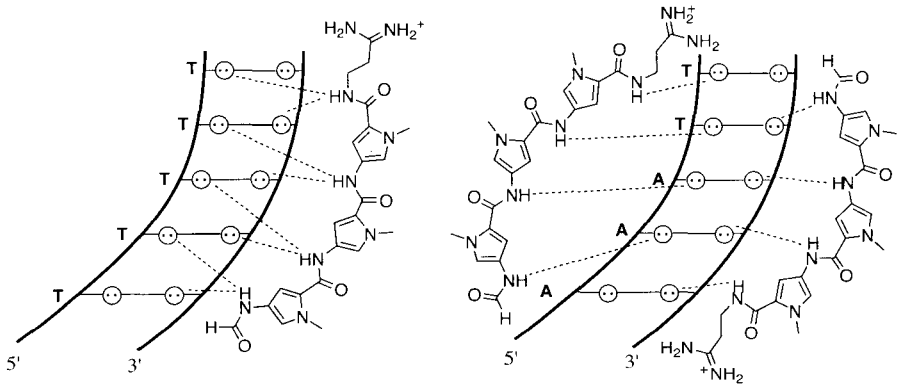


Figure 1.6. A schematic representation of binding in the minor groove by distamycin in a 1:1 and 2:1 complex.

hydrogen bond between 3-hydroxy of Hp and the second lone pair on the O2 of thymine, as well as a steric destabilization if Hp were placed over an adenine (Figure 1.7).

The affinity of 2-imidazole netropsin is modest (micromolar range). To increase its affinity, two molecules of 2-imidazole netropsin were connected in a head to tail fashion with an alkyl linker. The resulting hairpin structure bound DNA with a 100-fold increase in affinity.[12] Connecting the two distamycin like components allowed for the design of heterodimeric polyamides to target asymmetric sequences; for example, ImPyPy-γ-PyPyPyDp bound the sequence 5'-WGWWW-3'. The development of a solid phase methodology to synthesize Py-Im polyamides has greatly facilitated the examination of other motifs.[13] Extension of the binding site size has been achieved by addition of more rings to the hairpin or by covalently two hairpins to create a tandem structure (Figure 1.8).

Application of Py-Im Polyamides in transcription regulation

Py-Im polyamides have been shown to regulate gene expression.[14] In one case Py-Im polyamides were designed to inhibit the binding of TATA-binding protein (TBP) and LEF-1, two essential transcription factors that bind in the promoter region of HIV-1. Individually, each was shown to inhibit *in vitro* HIV-1 transcription between 60% and 80%. However, together the two polyamides showed synergistic effects and reduced HIV-1 transcription to undetectable levels in cell culture. This experiment showed that Py-Im polyamides are capable of inhibiting transcription when rationally designed to target specific sequences.

In a second case Py-Im polyamides are designed to activate instead of inhibit gene transcription.[15] The polyamide ImPyPyPy-γ-PyPyPyPy-β-Dp was coupled to a known transcriptional activator peptide. A promoter with six match sites for the polyamide-peptide conjugate was placed 40 base pairs upstream from the TATA box. Up to 40-fold activation of transcription was observed over basal levels with these conjugates. These

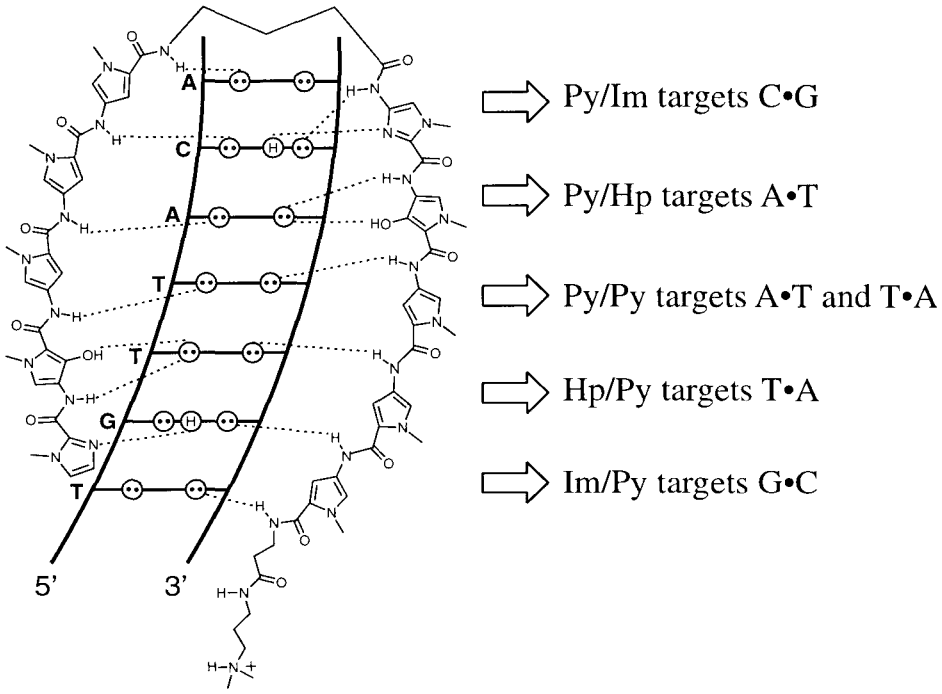


Figure 1.7. A schematic representation of the polyamide pair rules.

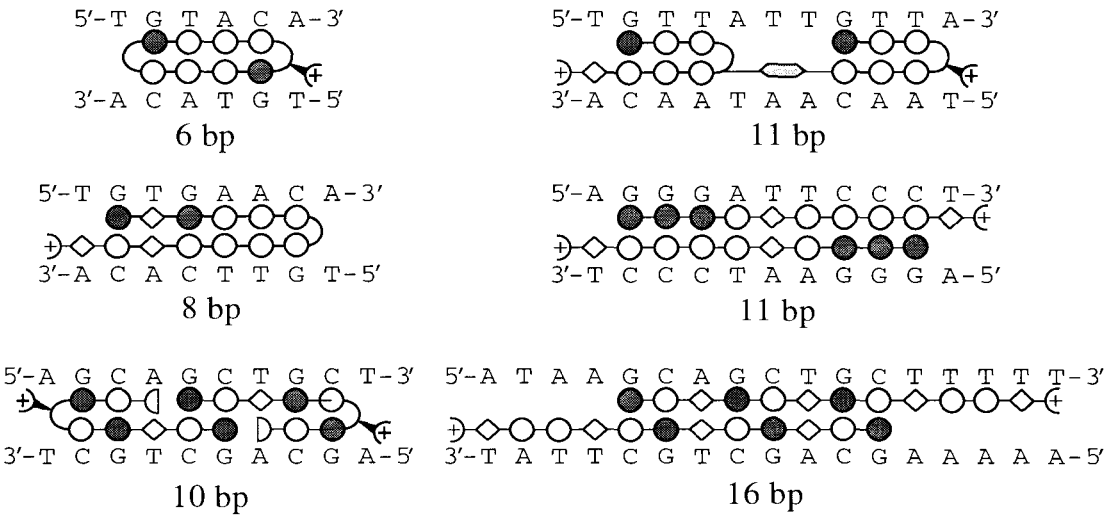


Figure 1.8. Representative motifs for polyamide:DNA recognition. Open circles and filled circles represent pyrrole and imidazole rings respectively. β-alanine and γ-turm are depicted as diamonds and curved lines respectively. The plus sign represents the dimethylaminopropylamide tail of the α-amino-γ-turn.

experiments show that polyamides can be used to manipulate a variety of transcriptional processes, and point the way to other experiments using small molecules that sequence specifically bind to DNA.

Scope of this Work

This first part of this thesis describes the development of new pairings at the terminus of polyamides to target a wider sequence variation with high affinity and specificities. A set of pairing rules has been developed to target all four base pairs. However, DNA recognition by several of these ring pairs can only be achieved when they are placed at an internal position in a hairpin rather than the terminal position in hairpin where "end effects" potentially alter molecular interactions necessary for specificity. Efforts to target T•A base pairs at the N-terminal position are described in chapters 2 and 3. Chapter 2 describes the synthesis and binding studies of hydroxypyrrole (Hp) ring for the targeting of T•A base pairs. Chapter 3 describes the development of a new ring hydroxybenzoic acid (Hb) in place of hydroxypyrrole for the recognition of T•A base pair at the N-terminus of polyamides. Efforts to target C•G base pairs using a pyrrole/imidazole are described in chapters 4 and 5. Efforts to target A•T base pairs are described in chapter 6. The second part of this thesis describes the design of polyamide conjugates to trap the endogenous enzyme Topoisomerase I. Chapter 7 describes the design and synthesis of a camptothecin polyamide conjugate capable of sequence specifically trapping Topoisomerase I. Chapter 8 describes the application of the camptothecin polyamide conjugate for the studies on gene repair *in vivo*.

References

1. Watson, J. D.; Crick, F. H. C. *Nature* **1953**, 171, 737.
2. Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. K.; Koopka, M. L. *Science* **1982**, 216, 475.

3. (a) Venter, C. et al. *Science* **2001**, 291, 1308. (b) International Human Genome Sequencing Consortium *Nature*, **2001**, 409, 860.

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

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

6. Krugh, T. R.; *Curr. Opin. Struct. Biol* **1994**, 4, 351-364.

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

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

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

10. (a) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111-115. (b) Kielkopf, C. L.; Bremer, R. B.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557-567.

11. (a) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471. (b) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 260-261. (c) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 11621-11629.

12. (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988 (b) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561. (c) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.

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

14. (a) Gottesfeld, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205. (b) Dickenson, L. A.; Gulizia, R.J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA*

1998, *95*, 12890-12895. (c) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Ghazal, P.; Dervan, P. B.; Gottesfeld, J. M. *Biochemistry*, **1999**, *38*, 10801-10807.

15. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000,** 97, 3930-3935.

Chapter 2

Recognition of T•A Sequences at the N-Terminal Position by Hairpin Polyamides with Hydroxypyrrole / Pyrrole Pairs

The text of this chapter was taken in part from a publication coauthored with Dr. Ulf Ellervik and Prof. Peter B. Dervan.
(Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, 122, 9354-9360.)

Abstract

In order to evaluate hairpin polyamides capable of distinguishing T•A/A•T base pairs at the terminal position, a series of hairpins with 3-hydroxy-1-methylpyrrole substituted at the 4-position by H, formamide and acetamide (Hp-**1**, Hp-**2**, Hp-**3**) were synthesized. The equilibrium association constants (K_a) were determined by quantitative DNase I footprint titration experiments for three pairs Hp-**1**/Py, Hp-**2**/Py, Hp-**3**/Py at four six base pair DNA sites, which differ at a common position 5'-T**N**TACA-3' (N=T, A, G, C). Remarkably only one pair Hp-**3**/Py afforded discrimination of T•A from A•T and with only modest sequence specificity. This study suggests that Hp/Py pairs may be limited within the hairpin motif to *internal* positions for T•A discrimination and there is a need for the design of new aromatic pairs for distinguishing T•A from A•T at the terminal position.

Introduction

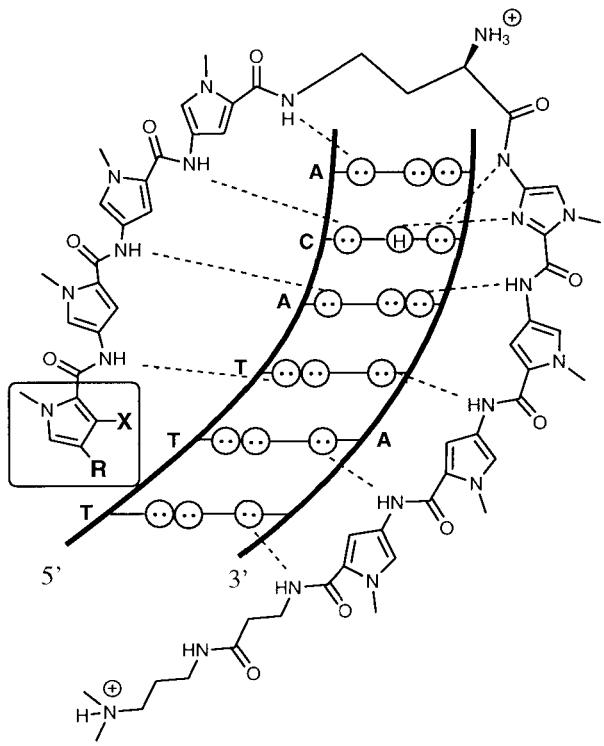
Polyamides containing three aromatic amino acids, N-methylpyrrole (Py), N-methylimidazole (Im) and 3-hydroxy-1-methylpyrrole (Hp), bind as antiparallel pairs in the minor groove of DNA with affinities and specificities for predetermined sequences comparable to DNA-binding proteins.[1] Im/Py pair binds G•C, while a Py/Im pair recognizes C•G. A Py/Py pair binds both A•T and T•A base pairs in preference to G•C and C•G. An Hp/Py pair prefers T•A over A•T, C•G or G•C. The T•A discrimination by Hp/Py pairs arises from a combination of specific hydrogen bonds with the thymine-O2 and shape selective recognition of an asymmetric cleft between the thymine-O2 and adenine-C2.[2]

Polyamide dimers linked by a γ turn unit creates a hairpin structure which sets the ring pairs unambiguously in register.[3] We have observed that hairpin polyamides with Im at the N-terminal position usually have higher affinities than the corresponding Py analogs.[1] The placement of an Im and Py in the first and eighth position of an eight-ring polyamide create an Im/Py pair at the N- and C-terminal positions of the hairpin, biasing our choice of target DNA sequences beginning in a G•C base pair.[1]

We would like to explore ring pairings that would allow the DNA sequence repertoire to be broadened wherein the target sequence begins with T or A. We have shown that Hp/Py pairs when placed *internal* in a hairpin distinguish T•A from A•T.[4] However, this has not been demonstrated for the terminal position in hairpin polyamides where "end effects" would potentially alter molecular interactions necessary for specificity. For example, the substituent at the 4-position of the N-terminal five-membered ring is flanked in the minor groove by β-alanine at the C-terminus. It is an important step forward in our gene regulation studies[5] to be able to target DNA sequences in promoters beginning with A•T or T•A base pairs.

We describe here a series of eight-ring polyamide hairpins with the general sequence XPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp with the N-terminal position X= 3-

hydroxy-1-methylpyrrole substituted at the 4-position with groups of increasing size H, formamide, acetamide (Hp-**1**, Hp-**2**, Hp-**3**) (Figure 2.1).



X	R = H	N(H)H	N(H)CH3
H	Py-1	Py-2	Py-3
OH	Hp-1	Hp-2	Hp-3

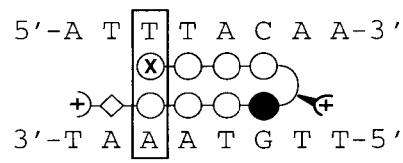


Figure 2.1. Hydrogen bonding model of the polyamide:DNA complex between eight-ring hairpin polyamides XPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (X = substituted Py or Hp) with 5'TTTACA-3' sites. A circle with two dots represents the lone pairs of N3 of purines and O2 of cytosine. Two touching circles with dots represent the two lone pairs of the O2 of thymine. Circles containing an H represent the exocyclic amino hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

Since Py is in the eighth position, the terminal pairs in the hairpin conformation are Hp-**1**/Py, Hp-**2**/Py and Hp-**3**/Py. For controls the parent pyrrole series without the 3-hydroxy was also synthesized, 1-methylpyrrole substituted at the 4-position with H, formamide and acetamide (Py-**1**, Py-**2**, Py-**3**) afford end cap pairs Py-**1**/Py, Py-**2**/Py and Py-**3**/Py (Figure 2.1). In addition, X= Im at the N-terminal was included to establish a baseline with previously published work. Seven eight-ring hairpin polyamides, differing in the terminal position, were synthesized for this study by solid-phase methods (Figure 2.2).[6]

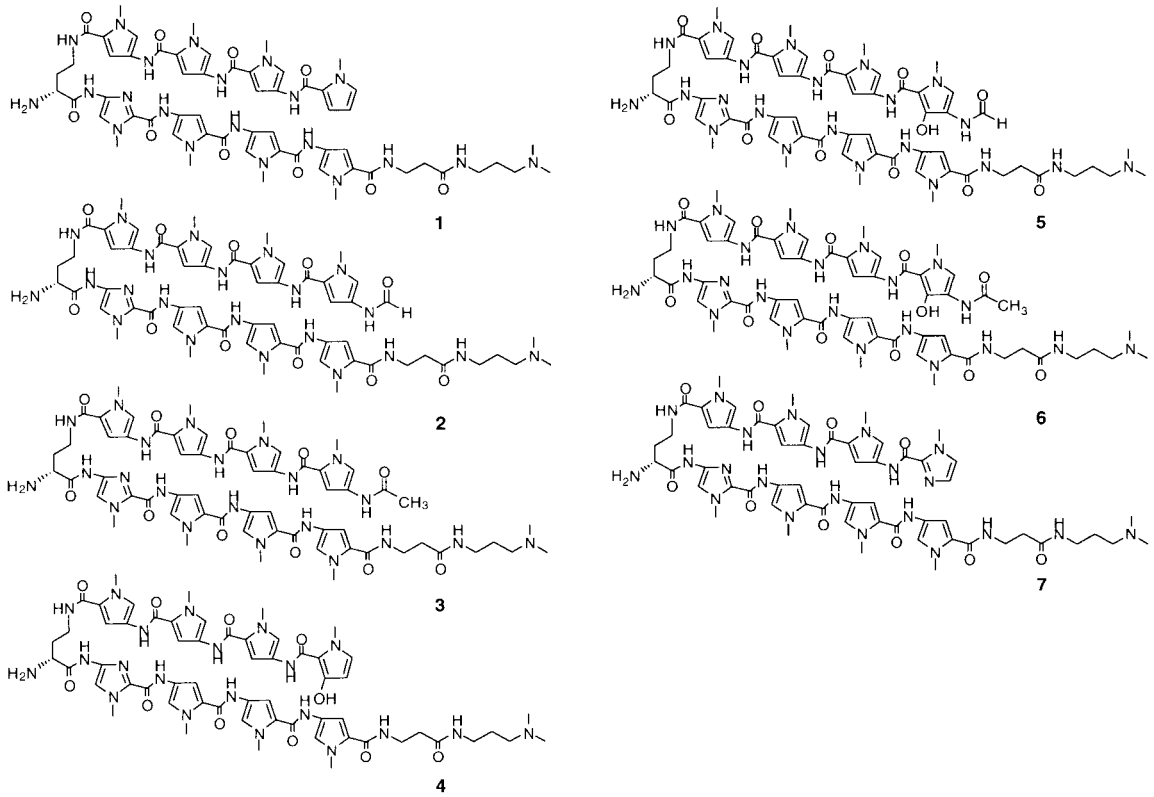


Figure 2.2. Structures of polyamides. (Py-**1**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**1**), (Py-**2**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**2**), (Py-**3**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**3**), (Hp-**1**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**4**), (Hp-**2**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**5**), (Hp-**3**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**6**), ImPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**7**).

The plasmid pCW15 was designed to contain the four six-base pair recognition sites 5'-TTTACA-3', 5'-TATACA-3', 5'-TGTACA-3' and 5'-TCTACA-3' which differ at a single common position allowing for comparison of the affinities between different terminal pairs and the four Watson Crick base pairs in the minor groove of DNA (Figure 2.3).

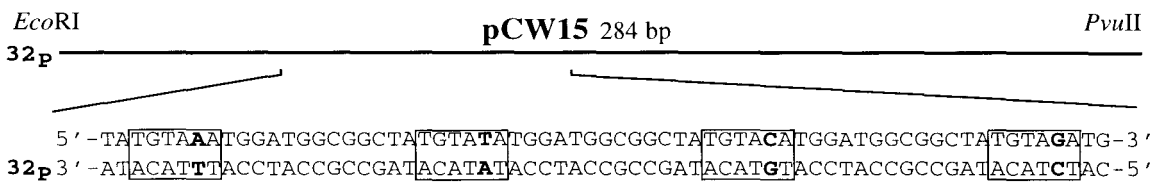


Figure 2.3. 284 base pair *Eco*RI/*Pvu*II restriction fragment derived from plasmid pCW15. The targeted six-base pair binding sites are shown in boxes.

Results

Synthesis. Initial attempts to synthesize 3-methoxy-1-methylpyrrole carboxylic acid (**13**) for incorporation in the polyamides were unsuccessful, probably due to the strong electrondonating properties of the methoxy group. The 3-benzoyloxy-1-methylpyrrole carboxylic acid **10** was synthesized, but all coupling attempts resulted in *N*-benzoylation of the polyamide-resin. Instead the 3-methoxy-1-methylpyrrole unit was introduced at the N-terminal position as a dimeric building block **11** which was synthesized in six steps from *t*-butyl 3-hydroxy-1-methylpyrrole-2-carboxylate (**8**) (Figure 2.4).[7] The hydroxy ester **8** was benzoylated using benzoyl chloride/pyridine to give compound **9**. The *t*-butyl ester was hydrolyzed using TMSOTf/DIEA[8] and pyrrole acid **10** was then coupled with an excess of methyl 4-amino-1-methylpyrrole-2-carboxylate.[6] Subsequent debenzoylation, methylation and hydrolysis of the methyl-ester afforded the dimeric building block **11**.

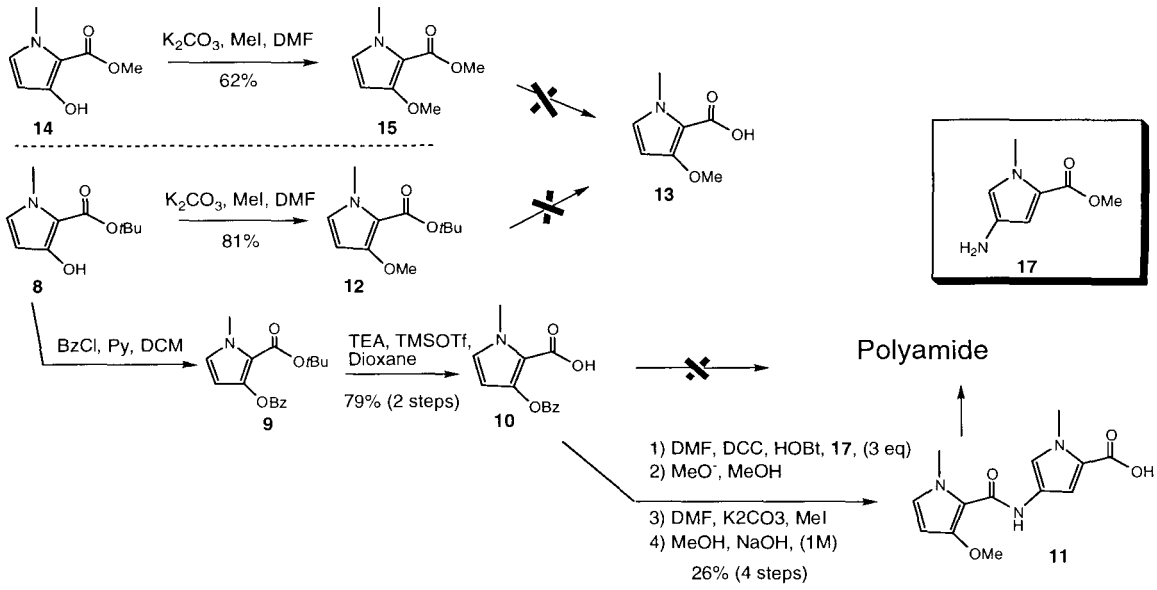


Figure 2.4. Synthesis of dimeric building block **11**.

The polyamide resins Boc-PyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin and Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin were synthesized in a stepwise manner from Boc-β-alanine-PAM resin (0.55 mmol/g) using Boc-chemistry solid-phase methodology (Figure 2.5).[6] The terminal monomers and dimer were then coupled to the deprotected resins. Hydroxypyrrole amino acid residues were introduced as orthogonally protected 3-methoxy-1-methylpyrrole (Op)[4c] derivatives. A sample of the resin was then cleaved by a single-step aminolysis reaction with ((dimethyl)amino)propylamine (55°C, 16 hours) and subsequently purified by reversed-phase HPLC to provide (Py-**1**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**1**), (Py-**2**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**2**), (Py-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**3**), and ImPyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**7**). Polyamides containing 3-methoxypyrrole at the N-terminus were deprotected by treatment with sodium ethanethiolate in DMF (100°C, 1 h) and purified by reversed-phase HPLC to provide (Hp-**1**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**4**), (Hp-**2**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**5**) and (Hp-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**6**).

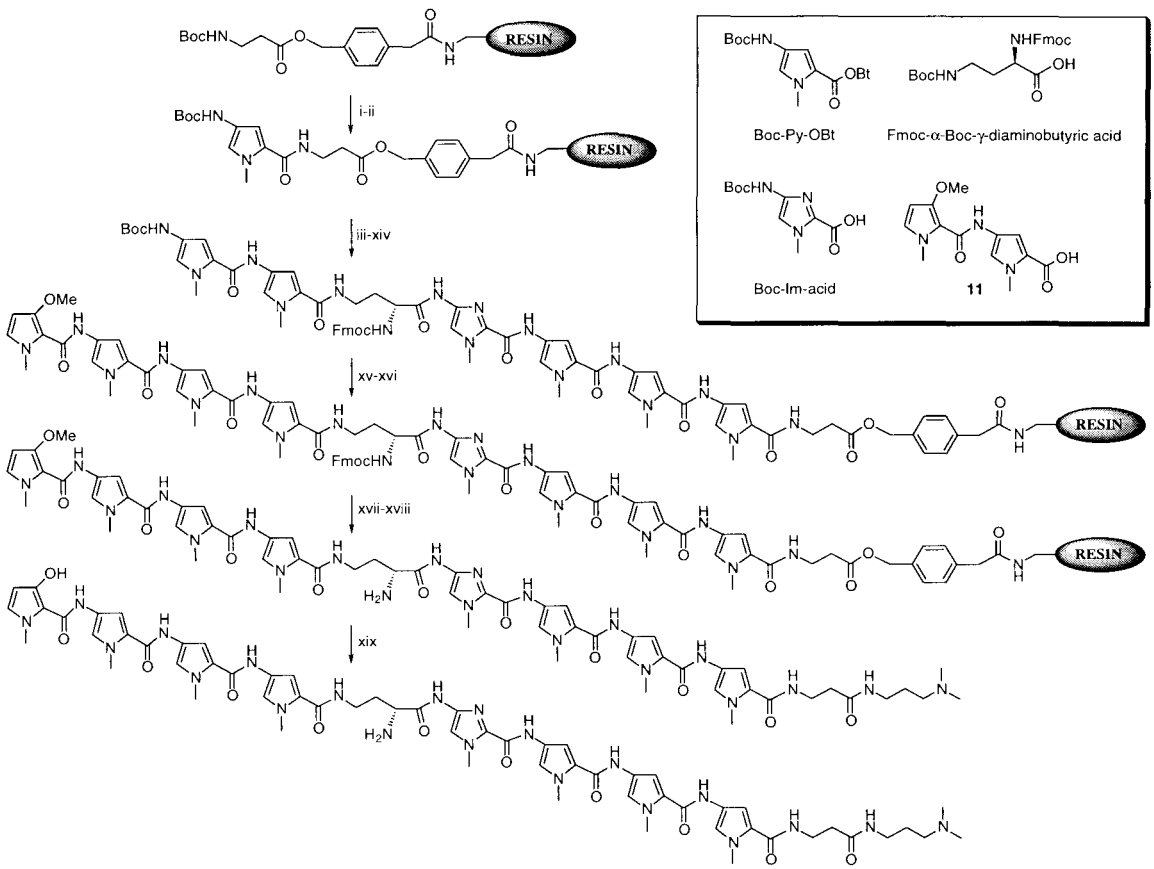


Figure 2.5. Solid-phase synthetic scheme for (Hp-**1**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**4**) starting from commercially available Boc-β-PAM-resin: (i) 80% TFA/DCM, 0.4 M PhSH. (ii) Boc-Py-OBt, DIEA, DMF. (iii) 80% TFA/DCM, 0.4 M PhSH. (iv) Boc-Py-OBt, DIEA, DMF. (v) 80% TFA/DCM, 0.4 M PhSH. (vi) Boc-Py-OBt, DIEA, DMF. (vii) 80% TFA/DCM, 0.4 M PhSH. (viii) Boc-Im-acid, DCC, HOBt, DIEA, DMF. (ix) 80% TFA/DCM, 0.4 M PhSH. (x) Fmoc-α-Boc-γ-diaminobutyric acid, HBTU, 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) **11**, DCC, HOBt, DIEA, DMF, r.t. 2h and then 37°C 2h. (xvii) Piperidine:DMF 3:1. (xviii) (Dimethylamino)propylamine, 55°C, 16h. (xix) NaH, EtSH, DMF, 100°C, 1 hour.

Quantitative DNase I footprinting titrations. Quantitative DNase I footprint titrations were performed on the 3'-^{32}P-end-labeled 284 base pair *Eco*RI/*Pvu*II restriction fragment of the pCW15 plasmid to determine the equilibrium association constant (K_a) of each eight-ring hairpin polyamide to the four different binding sites 5'-TNTACA-3' (where N=T, A, G, C)(Table 1, Figure 2.6). Polyamide **1**-**3**, with Py/Py pairs in the terminal position, bound to A•T and T•A with higher affinities than for G•C or C•G but did not distinguish T•A from A•T as expected. Remarkably, within the Hp/Py series, only polyamide **6** with the Hp-**3**/Py terminal pairs showed a preference for T•A over A•T.

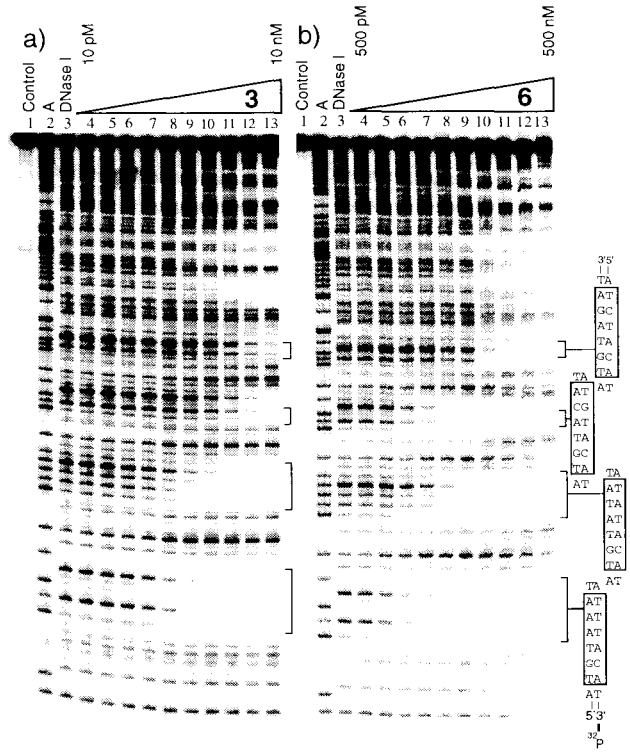


Figure 2.6. (a) Quantitative DNase I footprint titration experiment with (Py-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**3**) on the 284 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW15: lane 1, intact DNA; lane 2, A specific reaction; lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM (Py-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**3**); (b) Quantitative DNase I footprint titration experiment with (Hp-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**6**) on the 284 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW15: lane 1, intact DNA; lane 2, A specific reaction; lane 3, DNase I standard; lanes 4-13, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM (Hp-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**6**); The four sites 5'-T**T**TACA-3', 5'-T**A**TACA-3', 5'-T**G**TACA-3' and 5'-T**C**TACA-3' sites that were analyzed are shown on the right side of the gel.

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

Polyamide	5'-TTTACA-3'	5'-TATACA-3'	5'-TGTACA-3'	5'-TCTACA-3'
1	$3.6\ (\pm0.7) \times 10^{10}$	$3.0\ (\pm0.3) \times 10^{10}$	$8.4\ (\pm0.9) \times 10^{9}$	$3.7\ (\pm0.9) \times 10^{9}$
2	$5.6\ (\pm0.8) \times 10^{10}$	$1.2\ (\pm0.2) \times 10^{11}$	$1.5\ (\pm0.2) \times 10^{10}$	$9.3\ (\pm1.2) \times 10^{9}$
3	$1.1\ (\pm0.2) \times 10^{10}$	$9.4\ (\pm2.0) \times 10^{9}$	$1.9\ (\pm0.1) \times 10^{9}$	$1.3\ (\pm0.2) \times 10^{9}$
4	$1.3\ (\pm0.2) \times 10^{10}$	$1.1\ (\pm0.3) \times 10^{10}$	$2.7\ (\pm0.9) \times 10^{9}$	$3.0\ (\pm3.0) \times 10^{8}$
5	$1.3\ (\pm0.8) \times 10^{10}$	$1.1\ (\pm0.9) \times 10^{10}$	$1.2\ (\pm0.9) \times 10^{10}$	$1.8\ (\pm1.0) \times 10^{9}$
6	$2.1\ (\pm0.2) \times 10^{9}$	$6.2\ (\pm1.6) \times 10^{8}$	$1.1\ (\pm0.3) \times 10^{9}$	$1.5\ (\pm0.4) \times 10^{8}$
7	$2.0\ (\pm0.5) \times 10^{10}$	$1.1\ (\pm0.2) \times 10^{10}$	$2.9\ (\pm0.5) \times 10^{11}$	$2.9\ (\pm0.5) \times 10^{10}$

[a]Values reported are mean values from at least three DNase I footprint 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 $MgCl_2$, and 5 mM $CaCl_2$.

Discussion

Polyamide **1,** with a Py/Py pair in the terminal position, binds the sequences 5'-**TT**TACA-3' and 5'-**TA**TACA-3' with comparable affinity (3.6×10^{10} M^{-1} vs 3.0×10^{10} M^{-1}). Polyamide **1** binds 5'-**TG**TACA-3' and 5'-**TC**TACA-3' sequences with 4.3 and 9.7 times lower affinity. The discrimination of T•A /A•T over G•C / C•G at the N-terminal position is comparable to results shown when the Py/Py pair is placed in internal positions of the hairpin polyamide.[3b] The introduction of an acetamide group on the 1-methylpyrrole does not change the specificity for T•A /A•T over G•C / C•G but lowers the affinity for all four binding sites by a factor of 3, probably due to steric interactions of the methyl group with the floor of the minor groove or the adjacent β-alanine. Changing the acetamide group to formamide which reduces the steric bulk of the substituent restores the affinity.

For the 3-hydroxy series, polyamide **4** (Hp-**1**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp was expected to bind the sequence 5'-**TT**TACA-3' more tightly than 5'-**TA**TACA-3' due to the (Hp-**1**)/Py pair at the N-terminal position. However, no sequence preference was observed. One possible explanation could be rotation of the terminal hydroxypyrrole to form a hydrogen bond between the 3-hydroxyl hydrogen and the carbonyl oxygen. This

conformation would place the key hydroxyl group away from the minor groove (Figure 7). On the other hand polyamide **6** binds the sequence 5'-T**T**TACA-3' with higher affinity than 5'-T**A**TACA-3', a 3.4-fold specificity. Perhaps the amide at the 4-position forms an hydrogen bond with the thymine-O2 and thereby locks the orientation of the terminal hydroxypyrrole in the proper configuration (Figure 7). As in the Py/Py series the introduction of the acetamide group in polyamide **6** decreases the overall binding affinity. Replacing the acetamide with formamide in polyamide **5** restored the binding affinity but the specificity was lost. Both Hp-**2**/Py and Hp-**3**/Py pairs at the terminal position afforded poor sequence discrimination as demonstrated by the similar affinities at the sequences 5'-T**T**TACA-3' and 5'-T**G**TACA-3'. In conclusion Hp/Py pairs may be limited to internal positions for T•A discrimination and there is a need for the design of new pairs for the terminal position for distinguishing T•A from A•T (and both from G•C and C•G) in the minor groove of DNA.

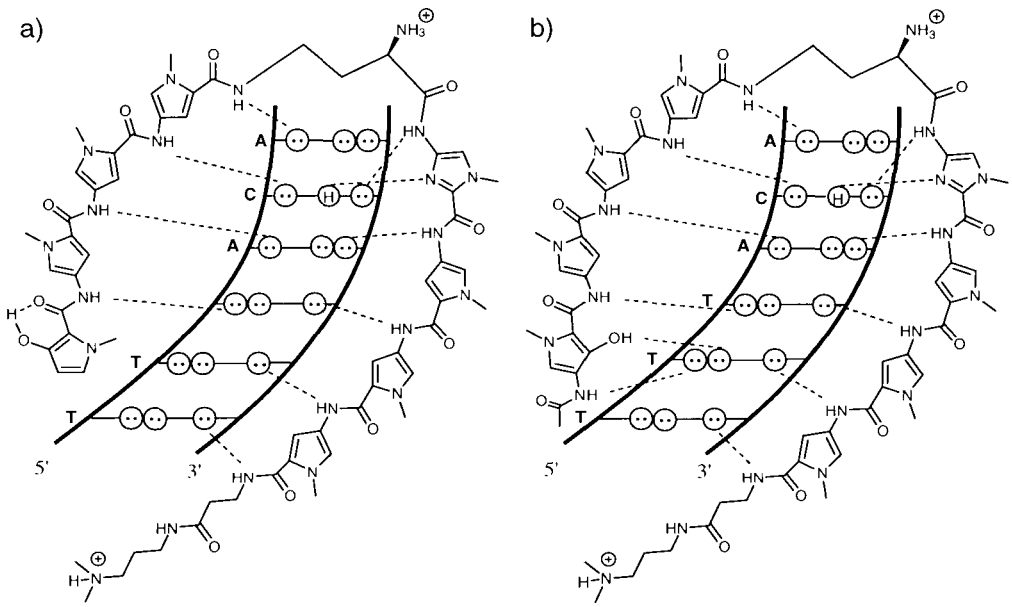


Figure 2.7. a) Binding model for the complex formed between (Hp-**1**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**4**), and a 5'-TTTACA-3' sequence. b) Binding model for the complex formed between (Hp-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**6**), and a 5'-TTTACA-3' sequence.

Experimental Section

Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-PAM-Resin), N,N'-Dicyclohexylcarbodiimide (DCC), N-Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA) and *N,N*-dimethylformamide (DMF) were purchased from Applied Biosystems. The DMF was distilled under reduced pressure prior to synthesis. (R)-2-Fmoc-4-Boc-diaminobutyric acid was from Bachem, dichloromethane (DCM) was reagent grade from EM, thiophenol (PhSH) and dimethylaminopropylamine from Aldrich, and trifluoroacetic acid (TFA) from Halocarbon. ^{1}H NMR were recorded on a Varian Mercury 300 instrument. Chemical shifts are reported in ppm downfield from the signal for Me_4Si, with reference to the solvent residual signal. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High-resolution fast atom bombardment mass spectra were recorded at the Mass Spectroscopy Laboratory at the University of California - Los Angeles. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5μm, 300 × 4.6 mm reversed-phase column in 0.1% (w/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 × 100 mm, 100 μm C_{18} column, 0.1% (w/v) TFA, 0.25% acetonitrile/min. 18MΩwater was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μm filtered. Thin-layer chromatography was performed on silica gel 60 F_{254} precoated plates. Reagent-grade chemicals were used unless otherwise stated.

***t*-Butyl 3-benzoyloxy-1-methylpyrrole-2-carboxylate (9).** *t*-Butyl 3-hydroxy-1-methylpyrrole-2-carboxylate[7] (**8**) (627 mg, 3.2 mmol) was dissolved in DCM (3 mL) and pyridine (0.515 mL, 6.4 mmol) was added. The mixture was cooled to 0°C and benzoyl chloride (0.553 mL, 4.8 mmol) was added during 5 minutes. The temperature was raised to room temperature during 1 hour. The mixture was partitioned between water and Et_2O and the water layer extracted with Et_2O two times. The combined organic phases were dried ($MgSO_4$) and concentrated. The crude product was purified by chromatography (SiO_2, 2:1 hexanes-EtOAc) to afford **9** (928 mg, 97%). ^1H NMR ($CDCl_3$) δ 8.10-8.25 (m, 2 H), 7.40-7.70 (m, 3 H), 6.68 (d, 1 H, J = 3.0 Hz), 6.02 (d, 1 H, J = 3.0 Hz), 3.88 (s, 3 H), 1.31 (s, 9 H); ^{13}C NMR ($CDCl_3$) δ 164.93, 160.07, 141.45, 134.72, 133.56, 130.75, 130.43, 129.89, 129.06, 128.60, 126.22, 102.16, 81.04, 38.35, 28.69. HRMS calcd. for $C_{17}H_{19}NO_4$: 301.1314; found: 301.1322.

3-Benzoyloxy-1-methylpyrrole-2-carboxylic acid (10). *t*-Butylester **9** (886 mg, 2.9 mmol) was dissolved in dioxane (19 mL) and DIEA (1.07 mL, 6.2 mmol) was added followed by TMSOTf (1.193 mL, 6.0 mmol) during 20 minutes under argon. The mixture was allowed to stir for 4 hours at room temperature and then partitioned between water and Et_2O and the water layer extracted with Et_2O two times. The combined organic phases were dried ($MgSO_4$) and concentrated. The crude product was purified by chromatography (SiO_2, 1:1 hexanes-EtOAc) to give **10** (438 mg, 61%) together with unreacted starting material (201 mg, 23%). ^1H NMR ($CDCl_3$) δ 8.10-8.30 (m, 2 H), 7.40-7.60 (m, 3 H), 6.77 (d, 1 H, J = 2.8 Hz), 6.20 (d, 1 H, J = 3.0 Hz), 3.88 (s, 3 H); ^{13}C NMR ($CDCl_3$) δ 164.99, 163.22, 143.27, 133.48, 133.21, 130.44, 130.06, 129.81, 128.61, 128.48, 127.23, 112.58, 102.42, 38.38 HRMS calcd. for $C_{13}H_{11}NO_4$: 245.0688; found: 245.0697.

4-[[[3-methoxy-1-methylpyrrole-2-yl]carbonyl]amino]-1-methylpyrrole-2-carboxylic acid (11). To a solution of **10** (438 mg, 1.8 mmol) in DMF (3 mL) was added DCC (361 mg, 1.8 mmol) and HOBt (229 mg, 1.7 mmol). The mixture was allowed to stir at

room temperature for 1.5 hours and the DCU was removed by filtration. Separately, to a solution of methyl 4-nitro-1-methylpyrrole-2-carboxylate[6] (1.00 g, 5.8 mmol) in DMF (4 mL) was added Pd/C (10%, 100 mg) and the mixture was hydrogenated in a Parr bomb apparatus (500 psi of H_2) for 1 hour. The catalyst was removed by filtration through Celite and the two solutions mixed and allowed to stir at 37°C for 6.5 hours and then at 55°C for 3 hours. The reaction mixture was poured on ice and extracted with Et_2O three times. The organic phase was dried ($MgSO_4$) and concentrated. The residue was dissolved in NaOMe-MeOH (0.05 M, 20mL) and allowed to stir at room temperature for 1.5 hours and then neutralized by Amberlite IR-120 H^+. The mixture was concentrated and the residue dissolved in DMF (10 mL) and K_2CO_3 (1.24 g, 9.0 mmol) was added followed by a solution of MeI (0.277 mL, 3.0 mmol in 1 mL of DMF) during 10 minutes. The mixture was allowed to stir at room temperature for 14 hours and then filtered and partitioned between water and DCM. The water phase was extracted with CH_2Cl_2 twice. The combined organic phases were dried ($MgSO_4$), concentrated and the ester was isolated by chromatography (SiO_2, 2:1 hexanes-EtOAc). The ester was dissolved in MeOH (10 mL), NaOH (aq, 1 M, 10 mL) was added, and the mixture was heated to 50°C for 20 h. All volatiles were removed *in vacuo* and the resulting solution was acidified by addition of HCl (aq, 3M). The solid was removed by filtration and dissolved in DCM/MeOH and subsequently purified by chromatography (SiO_2, 15:1 DCM/MeOH) to afford **11** (130 mg, 26%). ^1H NMR ($CDCl_3$) δ 8.74 (s, 1 H), 7.54 (d, 1 H, J = 1.8 Hz), 6.82 (d, 1 H, J = 2.1 Hz), 6.54 (d, 1 H, J = 2.7 Hz), 5.82 (d, 1 H, J = 3.0 Hz), 3.94, 3.92, 3.90 (s, 3 H each) ^{13}C NMR ($CDCl_3$); δ 165.60, 159.11, 150.63, 125.48, 122.65, 122.36, 119.04, 110.12, 109.78, 92.97, 58.54, 38.28, 37.28. HRMS calcd. for $C_{13}H_{15}N_3O_4$: 277.1063; found: 277.1052.

Boc-PyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin and **Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin.** Boc-PyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (0.350 mmol/g) and Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (0.335 mmol/g) were synthesized in a stepwise fashion from 0.55 mmol/g Boc-β-PAM-resin by manual solid-phase methods.[6] (R)-

2-Fmoc-4-Boc-diaminobutyric acid was incorporated as previously described for Boc-γ-aminobutyric acid.[3c,6]

Procedure for cleavage from the resin. After the coupling was completed the resin was filtered off the reaction and washed with DMF (2 × 30 s). DMF (1 mL) and piperidine (3 mL) were added and the mixture shaken for 30 minutes at 22°C. The resin was filtered off and washed with DMF (2 × 30 s), DCM (3 × 30 s), MeOH (1 × 30 s), Et_2O (1 × 30 s) and dried *in vacuo*. The resin was treated with (dimethylamino)propylamine (1 mL) with periodic agitation at 55°C for 16 hours. The reaction mixture was then filtered to remove the resin, TFA (0.1% (w/v), 7 mL) added and the resulting solution purified by reversed-phase HPLC. The pure compounds were recovered as white powders upon lyophilization of the appropriate fractions.

Deprotection of 3-methoxy-1-methylpyrrole polyamides. To a slurry of NaH (80 mg, 60% in mineral oil) in DMF (0.50 mL) was added a solution of ethanethiol (0.32 mL) in DMF (0.50 mL) and the mixture was heated to 100°C for 5 minutes in a sealed tube. The polyamide, dissolved in DMF (1.00 mL) was added to the ethanethiolate solution and the mixture was heated to 100°C for 1 hour in a sealed tube and then cooled to 0°C. The mixture was added to HOAc (3 mL) and all volatiles were removed (high vacuum, 40°C). The residue was dissolved in CH_3CN (1 mL) and TFA (7 mL, 0.1% (w/v)) and purified by preparative HPLC.

(Py-1)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (1). 1-methylpyrrole-2-carboxylic acid (50 mg, 0.40 mmol) was dissolved in DMF (1 mL) and HBTU (75 mg, 0.20 mmol) and DIEA (0.5 mL) were added and the mixture stirred for 10 minutes and added to deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (100 mg) and the mixture shaken for 4 hours at 22°C. The compound was then cleaved from the resin according to the general procedure. Preparative HPLC gave **1** (20 mg, 49% recovery). UV λ_{max} (H_2O) 240, 308 (69500); ^{1}H NMR (DMSO-d_6) δ 10.97, 10.06, 9.90, 9.89, 9.84, 9.78 (s, 1 H each), 8.28 (bs, 3 H), 8.13 (t, 1 H, J = 5.5 Hz), 7.99 (q, 1 H, J = 6.0 Hz), 7.48 (s, 1 H), 7.20 (d, 1 H, J = 1.5

Hz), 7.17 (t, 1 H, J = 1.8 Hz), 7.14 (d, 1 H, J = 1.5 Hz), 7.12 (t, 1 H, J = 1.8 Hz), 7.09 (d, 1 H, J = 1.8 Hz), 7.02 (s, 2 H), 6.99 (d, 1 H, J = 1.8 Hz), 6.91 (d, 1 H, J = 1.8 Hz), 6.89 (t, 1 H, J = 2.1 Hz), 6.85 (dd, 1 H, J = 3.9, 1.8 Hz), 6.82 (d, 1 H, J = 1.8 Hz), 5.99 (dd, 1 H, J = 3.9, 2.7 Hz), 3.92, 3.81, 3.80, 3.79, 3.78, 3.77, 3.75, 3.74 (s, 3 H each), 3.20-3.40 (m, 4 H), 3.05 (q, 2 H, J = 6.3 Hz), 2.90-3.00 (m, 2 H), 2.67 (d, 6 H, J = 4.5 Hz), 2.29 (t, 2 H, J = 7.2 Hz), 1.90-2.00 (m, 2 H), 1.68 (p, 2 H, J = 7.8 Hz); MALDI-TOF-MS calcd. for $C_{59}H_{73}N_{21}O_{10}$ (M + H): 1236.6; found: 1236.7.

(Py-2)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (2). Boc-Pyrrole-OBt ester[6] (50 mg, 0.14 mmol) was dissolved in DMF (0.5 mL) and DIEA (0.25 mL) and added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (50 mg) and the mixture shaken for 4 hours at 22°C. The resin was washed with DMF (2 × 30 s) and DCM (2 × 30 s) and the Boc group cleaved with 80% TFA/DCM/0.5 M PhSH (1 × 30 s, 1 × 20 min). The resin was washed with DCM (2 × 30 s) followed by DMF (2 × 30 s). Formic acid (1 mL) and Ac_2O (0.2 mL) were added and the mixture shaken for 1 h. The compound was then cleaved from the resin according to the general procedure. Preparative HPLC gave **2** (4.5 mg, 21% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 11.01, 10.10, 10.05 (s, 1 H each), 9.90-10.00 (m, 4 H), 9.88 (s, 1 H), 8.32 (bs, 3 H), 8.18 (t, 1 H, J = 5.7 Hz), 8.10 (d, 1 H, J = 1.8 Hz), 8.03 (q, 2 H, J = 5.7 Hz), 7.52 (s, 1 H), 7.24 (d, 1 H, J = 1.5 Hz), 7.20 (d, 2 H, J = 1.5 Hz), 7.19 (d, 1 H, J = 1.5 Hz), 7.15-7.18 (m, 3 H), 7.13 (d, 1 H, J = 1.5 Hz), 7.04-7.08 (m, 3 H), 6.96 (d, 1 H, J = 1.5 Hz), 6.90 (d, 1 H, J = 2.1 Hz), 6.87 (d, 1 H, J = 1.5 Hz), 3.96, 3.84 (s, 3 H each), 3.83, 3.82 (s, 6 H each), 3.79, 3.78 (s, 3 H each), 3.09 (q, 2 H, J = 5.4 Hz), 2.94-3.04 (m, 2 H), 2.75, 2.74 (s, 1 H each), 2.71 (d, 6 H, J = 4.8 Hz), 2.33 (t, 2 H, J = 7.5 Hz), 1.20-2.00 (m, 8 H). MALDI-TOF-MS calcd. for $C_{60}H_{74}N_{22}O_{11}$ (M + H): 1279.6; found: 1279.6.

(Py-3)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (3). Boc-Pyrrole-OBt ester[6] (100 mg, 0.28 mmol) was dissolved in DMF (1 mL) and DIEA (0.5 mL) and added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (100 mg) and the mixture shaken for 4 hours at

22°C. The resin was washed with DMF (2 × 30 s) and DCM (2 × 30 s) and the Boc group cleaved with 80% TFA/DCM/0.5 M PhSH (1 × 30 s, 1 × 20 minutes). The resin was washed with DCM (2 × 30 s) followed by DMF (2 × 30 s). DMF (1 mL) and Ac_2O (0.5 mL) were added and the mixture shaken for 5 minutes . The compound was then cleaved from the resin according to the general procedure. Preparative HPLC gave **3** (15 mg, 35% recovery). UV λ_{max} (H_2O) 240, 308 (69500); ^1H NMR (DMSO-d_6) δ 10.97, 10.07, 9.90 (s, 1 H each), 9.88 (s, 2 H), 9.85, 9.84, 9.78 (s, 1 H each), 8.27 (bs, 3 H), 8.13 (t, 1 H, *J* = 4.8 Hz), 8.00 (q, 2 H, *J* = 6.0 Hz), 7.48 (s, 1 H), 7.21 (d, 1 H, *J* = 1.5 Hz), 7.16 (t, 1 H, *J* = 1.2 Hz), 7.14 (d, 1 H, *J* = 1.5 Hz), 7.12 (bs, 1 H), 7.08 (dd, 1 H, *J* = 3.0, 1.2 Hz), 7.02 (d, 1 H, *J* = 1.8 Hz), 7.01 (d, 1 H, *J* = 1.8 Hz), 6.91 (d, 1 H, *J* = 1.8 Hz), 6.82 (d, 1 H, *J* = 1.5 Hz), 6.80 (d, 1 H, *J* = 2.1 Hz), 3.92, 3.79 (s, 3 H each), 3.78 (s, 6 H), 3.77, 3.76, 3.75, 3.74 (s, 3 H each), 3.20-3.35 (m, 4 H), 3.04 (q, 2 H, *J* = 6.6 Hz), 2.94 (p, 2 H, *J* = 6.6 Hz), 2.28 (t, 2 H, *J* = 7.5 Hz), 1.91 (s, 3 H), 1.67 (p, 2 H, *J* = 7.5 Hz). MALDI-TOF-MS calcd. for $C_{61}H_{76}N_{22}O_{11}$ (M + H): 1293.6; found: 1293.7.

(Hp-1)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (4). Compound **11** (22 mg, 0.08 mmol) was dissolved in DMF (0.3 mL) and HOBt (9 mg, 0.07 mmol) and DCC (15 mg, 0.07 mmol) were added and the mixture was allowed to stir for 20 minutes and then added to the deprotected Boc-PyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (50 mg) followed by DIEA (0.1 mL) and the mixture shaken for 2.0 hours at 22°C and then 2.0 hours at 37°C. The compound was then cleaved from the resin according to the general procedure. Preparative HPLC gave methyl protected (Hp-1)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (9 mg, 39% recovery).UV λ_{max} (H_2O) 240, 312 (69500); ^1H NMR (DMSO-d_6) δ 10.97, 10.06, 9.90, 9.89, 9.84, 9.82, 8.73 (s, 1 H each), 8.27 (bs, 3 H), 8.13 (t, 1 H, *J* = 5.7 Hz), 8.00 (q, 2 H, *J* = 5.7 Hz), 7.48 (s, 1 H), 7.23 (d, 1 H, *J* = 1.8 Hz), 7.20 (d, 1 H, *J* = 1. 8 Hz), 7.15 (bs, 1 H), 7.12 (bs, 1 H), 7.09 (d, 1 H, *J* = 1. 8 Hz), 7.02 (bs, 2 H), 6.93 (d, 1 H, *J* = 1.8 Hz), 6.91 (d, 1 H, *J* = 1.8 Hz), 6.82 (d, 1 H, *J* = 1.8 Hz), 6.80 (d, 1 H, *J* = 2.7 Hz), 5.90 (d, 1 H, *J* = 2.7 Hz), 3.92, 3.80, 3.79 (s, 3 H each), 3.78 (bs, 9 H), 3.76, 3.75, 3.74 (s, 3 H each), 3.04 (q, 2

H, J = 6.0 Hz), 2.94 (p, 2 H, J = 5. 0 Hz). 2.67 (d, 6 H, J = 4.5 Hz), 2.28 (t, 2 H, J = 7.2 Hz), 1.92 (bs, 2 H), 1.67 (p, 2 H, J = 7.2 Hz). MALDI-TOF-MS calcd. for $C_{60}H_{75}N_{21}O_{11}$ (M + H): 1266.6; found: 1266.7. The methyl protected (Hp-1)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (2.1 mg, 1.7 μmol) was deprotected to give **4** (1.3 mg, 63% recovery). UV λ_{max} (H_2O) 240, 310 (69500); ^{1}H NMR (DMSO-d_6) δ 11.12, 10.44, 10.19, 10.03, 10.02, 9.96, 9.89, 9.09 (s, 1 H each), 8.39 (bs, 3 H), 8.26 (t, 1 H, J = 2.2 Hz), 8.12 (q, 2 H, J = 2.1 Hz), 7.61 (s, 1 H), 7.36 (d, 1 H, J = 1.5 Hz), 7.33 (d, 1 H, J = 1.5 Hz), 7.28 (d, 2 H, J = 1.5 Hz), 7.25 (bd, 2 H, J = 1.8 Hz), 7.20 (d, 1 H, J = 1.5 Hz), 7.12-7.18 (m, 2 H), 7.05 (t, 2 H, J = 2.1 Hz), 6.96 (d, 1 H, J = 1.5 Hz), 6.80 (d, 1 H, J = 2.1 Hz), 5.73 (d, 1 H, J = 2.7 Hz), 4.05, 3.93, 3.92, 3.91, 3.90, 3.88 (s, 3 H each), 3.87 (s , 6 H), 3.17 (q, 2 H, J = 6.0 Hz), 3.06 (bt, 2 H, J = 8.1 Hz), 2.80 (s, 3 H), 2.42 (t, 2 H, J = 7.2 Hz), 2.06 (bs, 2 H), 1.80 (bt, 2 H, J = 7.8 Hz), 1.30 (s, 1 H); MALDI-TOF-MS calcd. for $C_{59}H_{73}N_{21}O_{11}$ (M + H): 1252.6; found: 1252.6.

(Hp-2)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (5). Boc-Op acid[4c] (50 mg, 0.18 mmol) was dissolved in DMF (1.0 mL) and HBTU (60 mg, 0.16 mmol) and DIEA (0.5 mL) were added and the mixture stirred for 10 minutes and added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (50 mg) and the mixture shaken for 4.0 hours at 22°C. The resin was washed with DMF (2 × 30 s) and DCM (2 × 30 s) and the Boc group cleaved with 80% TFA/DCM/0.5 M PhSH (1 × 30 s, 1 × 20 minutes). The resin was washed with DCM (2 × 30 s) followed by DMF (2 × 30 s). Formic acid (1 mL) and Ac_2O (0.2 mL) were added and the mixture shaken for 1 h. The compound was then cleaved from the resin according to the general procedure. Preparative HPLC gave methyl protected (Hp-2)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (6 mg, 27% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 11.10, 10.19, 10.03, 10.02, 9.99, 9.96, 9.88, 9.21 (s, 1 H each), 8.41 (bs, 3 H), 8.27 (t, 1 H, J = 6.0 Hz), 8.23 (d, 1 H, J = 1.5 Hz), 8.12 (q, 2 H, J = 6.0 Hz), 7.61 (s, 1 H), 7.36 (d, 1 H, J = 1.8 Hz), 7.33 (d, 1 H, J = 1.5 Hz), 7.29 (s, 1 H), 7.28 (s, 2 H), 7.25 (s, 2 H), 7.22 (d, 1 H, J = 1.5 Hz), 7.15 (d, 2 H, J = 1.8 Hz), 7.14 (d, 1 H, J = 1.5 Hz), 7.04 (d, 1 H, J = 1.5 Hz), 6.96 (d, 1 H, J = 1.5 Hz), 4.05 (s, 3 H), 3.93 (s, 6 H),

3.91, 3.90, 3.88,3.87, 3.86, 3.85 (s, 3 H each), 3.17 (q, 2 H, J = 6.3 Hz), 3.07 (p, 2 H, J = 5.4 Hz), 2.80 (d, 6 H, J = 4.5 Hz), 2.42 (t, 2 H, J = 7.5 Hz), 2.02-2.12 (m, 2 H), 1.81 p, 2 H, J = 7.8 Hz); MALDI-TOF-MS calcd. for $C_{61}H_{76}N_{22}O_{12}$ (M + H): 1309.6; found: 1309.7. The methyl protected (Hp-2)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (1.8 mg, 1.4 μmol) was deprotected according to the general procedure to give **5** (0.7 mg, 37% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 10.96, 10.60, 10.21, 10.06 (s, 1 H each), 9.89 (bs, 2 H), 9.83, 9.79 (s, 1 H each), 8.99 (s, 1 H), 8.25 (bs, 3 H), 8.13 (bs, 1 H), 7.99 (bs, 3 H), 7.48, 7.24, 7.20 (s, 1 H each), 7.14, 7.11 (s, 2 H each), 7.08, 7.02, 7.00, 6.94, 6.91 (s, 1 H each), 6.83 (s, 2 H), 3.91 (s, 3 H), 3.78, 3.77, 3.75, (s, 3 H each), 3.73 (s, 6 H), 3.04 (q, 2 H, J = 6.6 Hz), 2.95 (p, 2 H, J = 6.6 Hz), 2.67 (d, 6 H, J = 3.9 Hz), 2.28 (t, 2 H, J = 7.2 Hz), 1.90-2.00 (m, 2 H), 1.67 (p, 2 H, 6.8 Hz), 1.17 (s, 1 H); MALDI-TOF-MS calcd. for $C_{60}H_{74}N_{22}O_{12}$ (M + H): 1295.6; found: 1295.8.

(Hp-3)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (6). Boc-Op acid[4c] (50 mg, 0.18 mmol) was dissolved in DMF (1.0 mL) and HBTU (60 mg, 0.16 mmol) and DIEA (0.5 mL) were added and the mixture stirred for 10 minutes and added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (100 mg) and the mixture shaken for 4.0 hours at 22°C. The compound was then cleaved from the resin according to the general procedure. Preparative HPLC gave methyl protected (Hp-3)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (15 mg, 46% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 10.99, 10.20, 9.92, 9.90 (s, 1 H each), 9.85 (s, 2 H), 9.34, 9.02 (s, 1 H each), 8.38 (bs, 3 H), 8.17 (t, 1 H, J = 5.9 Hz), 7.96-8.10 (m, 2 H), 7.47 (s, 1 H), 7.23 (d, 1 H, J = 1.2 Hz), 7.21 (bs, 1 H), 7.17 (d, 1 H, J = 1.2 Hz), 7.16 (bs, 2 H), 7.13 (d, 1 H, J = 1.2 Hz), 7.09-7.12 (m, 4 H), 7.05 (s, 1 H), 7.02 (bs, 2 H), 6.99 (d, 1 H, J = 1.8 Hz), 6.90 (d, 1 H, J = 1.2 Hz), 6.81 (bs, 2 H), 3.92 (s, 3 H), 3.89 (s, 1 H), 3.79, 3.78 (s, 6 H each), 3.75 (s, 3 H each), 3.73 (s, 6 H), 3.71 (s, 3 H), 3.20-3.35 (m, 4 H), 3.04 (q, 2 H, J = 6.3 Hz), 2.93 (p, 2 H, J = 4.5 Hz), 2.65 (d, 6 H, J = 4.5 Hz), 2.28 (t, 2 H, J = 6.6 Hz), 1.96 (s, 1 H), 1.95 (s, 3 H), 1.69 (p, 2 H, J = 7.2 Hz). MALDI-TOF-MS calcd. for $C_{62}H_{78}N_{22}O_{12}$ (M + H): 1323.6; found: 1323.8. The methyl

protected (Hp-3)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (3.1 mg, 2.4 μmol) was deprotected according to the general procedure to give **6** (0.7 mg, 23% recovery). UV λ_{max} (H_2O) 240, 308 (69500); ^{1}H NMR (DMSO-d_6) δ 11.12, 10.97, 10.17, 10.06, 9.92, 9.89, 9.84, 9.78 (s, 1 H each), 9.24 (bs, 1 H), 8.26 (bs, 3 H), 8.13 (t, 1 H, J = 5.6 Hz), 8.00 (q, 2 H, J = 5.67 Hz), 7.48, 7.24, 7.20 (s, 1 H each), 7.15, 7.11 (s, 2 H each), 7.09, 7.02, 7.00, 6.93, 6.91, 6.83, 6.76 (s, 1 H each), 3.91, 3.79 (s, 3 H each), 3.78 (s, 6 H), 3.77, 3.74, 3.73, 3.72 (s, 3 H each), 3.04 (q, 2 H, J = 6.0 Hz), 2.94 (p, 2 H, J = 3.9 Hz), 2.68 (d, 6 H, J = 3.3 Hz), 2.28 (t, 2 H, J = 6.6 Hz), 1.99 (s, 3 H), 1.94 (bs, 2 H), 1.67 (p, 2 H, J = 5.8 Hz), 1.17 (s, 1 H); MALDI-TOF-MS calcd. for $C_{61}H_{76}N_{22}O_{12}$ (M + H): 1309.6; found: 1309.7.

ImPyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (7). Imidazole acid (50 mg, 0.40 mmol) was dissolved in DMF (1.0 mL) and HBTU (140 mg, 0.37 mmol) and DIEA (0.5 mL) were added and the mixture stirred for 10 minutes and added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (100 mg) and the mixture shaken for 2.5 hours at 22°C. Preparative HPLC gave **7** (20 mg, 50% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 10.97, 10.42, 10.06 (s, 1 H each), 9.91 (bs, 2 H), 9.88, 9.84 (s, 1 H each), 8.28 (bs, 2 H), 8.13 (t, 1 H, J = 5.7 Hz), 7.99 (q, 2 H, J = 5.7 Hz), 7.48, 7.34 (s, 1 H each), 7.23 (d, 1 H, J = 1.6 Hz), 7.20 (d, 1 H, J = 1.4 Hz), 7.17 (d, 1 H, J = 1.4 Hz) 7.15 (bs, 1 H), 7.10-7.20 (m, 2 H), 7.09 (bs, 1 H), 7.02 (bs, 2 H), 6.99 (t, 1 H, J = 1.2 Hz), 6.91 (bs, 1 H), 6.83 (bs, 1 H), 3.93 (s, 6 H), 3.92, 3.79, 3.78, 3.77, 3.75, 3.73 (s, 3 H each), 3.31 (q, 2 H, J = 5.7 Hz), 3.23 (q, 2 H, J = 5.7 Hz), 3.04 (q, 2 H, J = 6.0 Hz), 2.90-2.97 (m, 2 H), 2.68 (d, 6 H, J = 4.8 Hz), 2.28 (t, 2 H, J = 7.2 Hz), 1.88-2.00 (m, 2 H), 1.67 (p, 2 H, J = 8.1 Hz). MALDI-TOF-MS calcd. for $C_{58}H_{72}N_{22}O_{10}$ (M + H): 1237.6; found: 1237.7.

Construction of plasmid DNA. The plasmid pCW15 was constructed by hybridization of the inserts, 5′-CTAGCGGCTATGTAAATGGATGGCGGCTATGTA TATGGATGGCGGCTATGTACATGGATGGCGGCTATGTAGATGGATTGCA-3′ and 5′-ATCCATCTACATAGCCGCCATCCATGTACATAGCCGCCATCCATATA

CATAGCCGCCATCCATTTACATAGCC-3′. The hybridized insert was ligated into linearized pUC19 *Xba*I /*Pst*I plasmid using T4 DNA ligase. The resultant constructs were used to transform Top10F′ OneShot competent cells from Invitrogen. Ampicillin-resistant white colonies were selected from 25 mL Luria-Bertani medium agar plates containing 50 μg/mL ampicillin and treated with XGAL and IPTG solutions. Large-scale plasmid purification was performed with Qiagen Maxi purification kits. Dideoxy sequencing was used to verify the presence of the desired insert.

Preparation of 3′- and 5′-End-Labeled Restriction Fragments. The plasmid pCW15 was linearized with *Eco*RI and *Pvu*II and then treated with Sequenase (version 2.0 from United States Biochemical), deoxyadenosine 5′-[α-^{32}P]triphosphate and thymidine 5′-[α-^{32}P]triphosphate for 3′ labeling. The 3' labeled fragment was loaded onto a 7% non-denaturing polyacrylamide gel, and the desired 284 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.[9]

Quantitative DNase I Footprinting. DNase I Footprinting reactions were carried out as previously described.[3b] Photostimulable storage phosphorimaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22°C for 12-16 hours. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integration of all bands using ImageQuant v. 3.2 software. All DNA manipulations were performed according to standard protocols.[10]

Acknowledgement

We are grateful to the National Institutes of Health (GM-27681) for research support, the Knut and Alice Wallenberg Foundation for a post-doctoral fellowship to U.E. and the

National Institutes of Health for a research traineeship award (GM-08501) to C.C.C.W. We thank G. M. Hathaway for MALDI-TOF mass spectrometry.

References

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

2. (a) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111-115. (b) Kielkopf, C. L.; Bremer, R. B.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557-567.

3. (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988 (b) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561. (c) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.

4. (a) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471. (b) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 260-261. (c) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 11621-11629.

5. (a) Gottesfeld, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205. (b) Dickenson, L. A.; Gulizia, R.J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895. (c) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Ghazal, P.; Dervan, P. B.; Gottesfeld, J. M. *Biochemistry*, **1999**, *38*, 10801-10807.

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

7. Wasserman, H. H.; Cook, J. D.; Fukuyama, J. M.; Rotello, V. M. *Tetrahedron Lett.* **1989**, *30*, 1721-1724.

8. Emde, H.; Domsch, D.; Feger, H.; Frick, U.; Götz, A.; Hergott, H. H.; Hofmann, K.; Kober, W.; Krägeloh, K.; Oesterle, T.; Steppan, W.; West, W.; Simchen, G. *Synthesis* **1982**, 1-26.

9. (a) Iverson, B. L.; Dervan, P. B. *Nucl. Acids Res.* **1987**, *15*, 7823-7830. (b) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499-560.

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

Chapter 3

Hydroxybenzamide/Pyrrole Pairs Distinguishes T•A from A•T Base Pairs in the Minor Groove of DNA

The text of this chapter was taken in part from a publication coauthored with Dr. Ulf Ellervik and Prof. Peter B. Dervan.
(Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, 122, 9354-9360.)

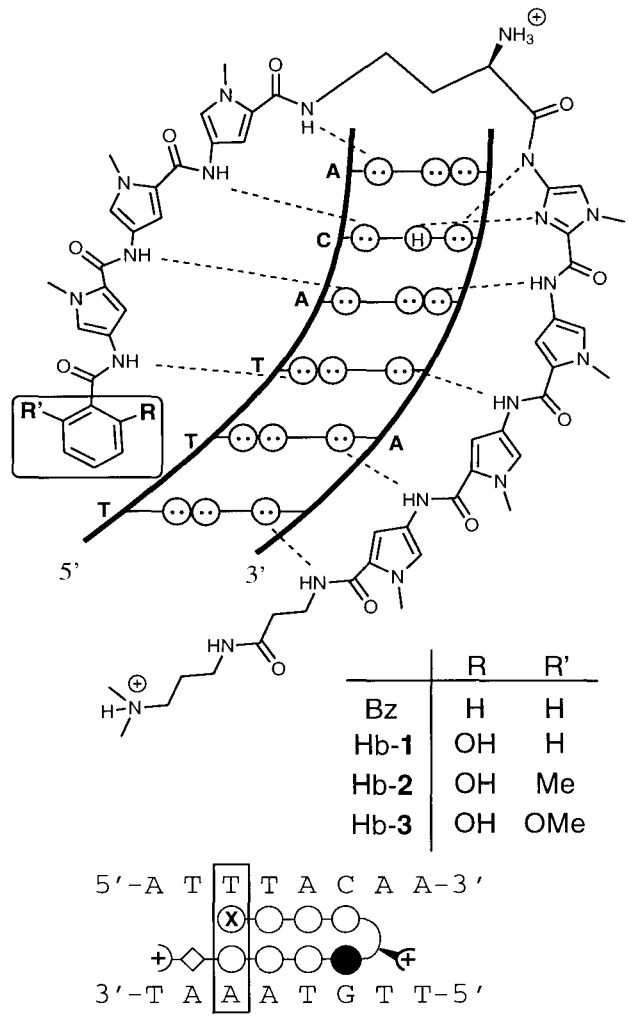
Abstract

A new aromatic pair, 2-hydroxy-6-methylbenzamide/1-methylpyrrole at the terminal position of hairpin polyamides has been designed for distinguishing T•A from A•T base pairs and both from G•C/C•G in the minor groove of DNA. Four eight-ring hairpin polyamides with benzamide (Bz), 2-hydroxybenzamide (Hb-**1**), 2-hydroxy-6-methylbenzamide (Hb-**2**), and 2-hydroxy-6-methylbenzamide (Hb-**3**) at the N-terminal position were synthesized. The equilibrium association constants (K_a) were determined at four DNA sites which differ at a single common position, 5'-T**N**TACA-3' (N = T, A, G, C). Quantitative DNase I footprint titration experiments reveal that (Hb-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**4**) bound the sequences 5'-T**T**TACA-3' and 5'-T**A**TACA-3' with high affinity $K_a = 2.6 \times 10^{10}$ M^{-1} and $K_a = 8.4 \times 10^{9}$ M^{-1}, respectively, a three-fold specificity for T vs A. Importantly, the sequences 5'-T**G**TACA-3' and 5'-T**C**TACA-3' are bound with 50 and 200 times lower affinity, revealing an overall specificity of Hb-**3**/Py of T>A>>G>C. These results expand the repertoire of sequences targetable by hairpin polyamides.

Introduction

The ability to distinguish T•A from A•T base pairs by synthetic ligands which bind in the minor groove of DNA is an essential milestone for targeting specific A/T rich sequences, within promoters for gene regulation studies.[1] We recently reported that the 2-hydroxypyrrole/pyrrole pair (Hp/Py)[2] for distinguishing T•A from A•T is most suitable for internal positions in polyamides.[3] The Hp ring at the N-terminal position affords lower affinity and sequence specificity for A•T/T•A vs G•C/C•G than anticipated.[3] During the early development of minor groove binders, it was shown that six-membered aromatic rings such as 2-pyridine, as well as the five membered ring imidazole Im, when paired with Py could target G•C base pairs.[4] The question arises whether 2-hydroxybenzene (Hb) at the N-terminal position of hairpin polyamides could act as a mimic of the 2-hydroxypyrrole (Hp) with improved affinity and specificity for T.

We describe here a series of eight-ring polyamide hairpins with the general sequence **X**PyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp and the N-terminal position X= benzamide (Bz), 2-hydroxybenzamide (Hb-**1**), 2-hydroxy-6-methylbenzamide (Hb-**2**), and 2-hydroxy-6-methoxybenzamide (Hb-**3**). Since Py is in the eighth position, the terminal pairs in the hairpin conformation are Hb-**1**/Py, Hb-**2**/Py and Hb-**3**/Py (Figure 3.1).
For controls the parent benzamide without the hydroxy is included. A total of 11 eight-ring hairpin polyamides differing in the terminal position were synthesized by solid phase methods (Figure 3.2a and b). The plasmid pCW15 was designed to contain four six-base pair recognition sites 5-T**N**TACA-3' (where N=T,A,G,C) which differ at a single common position allowing for comparison of the affinities between different terminal pairs and the four Watson Crick base pairs in the minor groove of DNA (Figure 3.3).



	R	R'
Bz	H	H
Hb-1	OH	H
Hb-2	OH	Me
Hb-3	OH	OMe

Figure 3.1. Hydrogen bonding model of the polyamide:DNA complex between eight-ring hairpin polyamides containing substituted benzamides with the 5'TTTACA-site. A circle with two dots represents the lone pairs of N3 of purines and O2 of cytosine. Two touching circles with dots represent the two lone pairs of the O2 of thymine. Circles containing an H represent the exocyclic amino hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

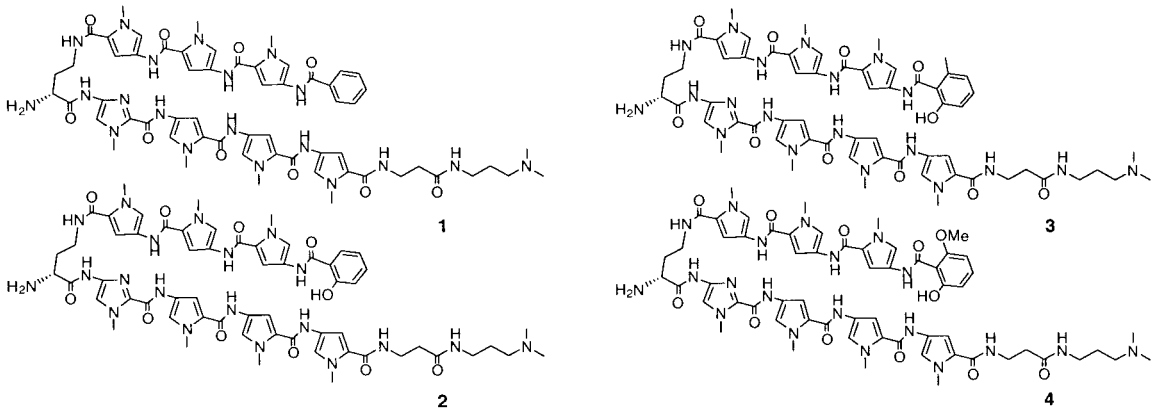


Figure 3.2a. Structures of polyamides **1**-**4**. BzPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**1**), (Hb-**1**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**2**), (Hb-**2**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**3**), (Hb-**3**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**4**).

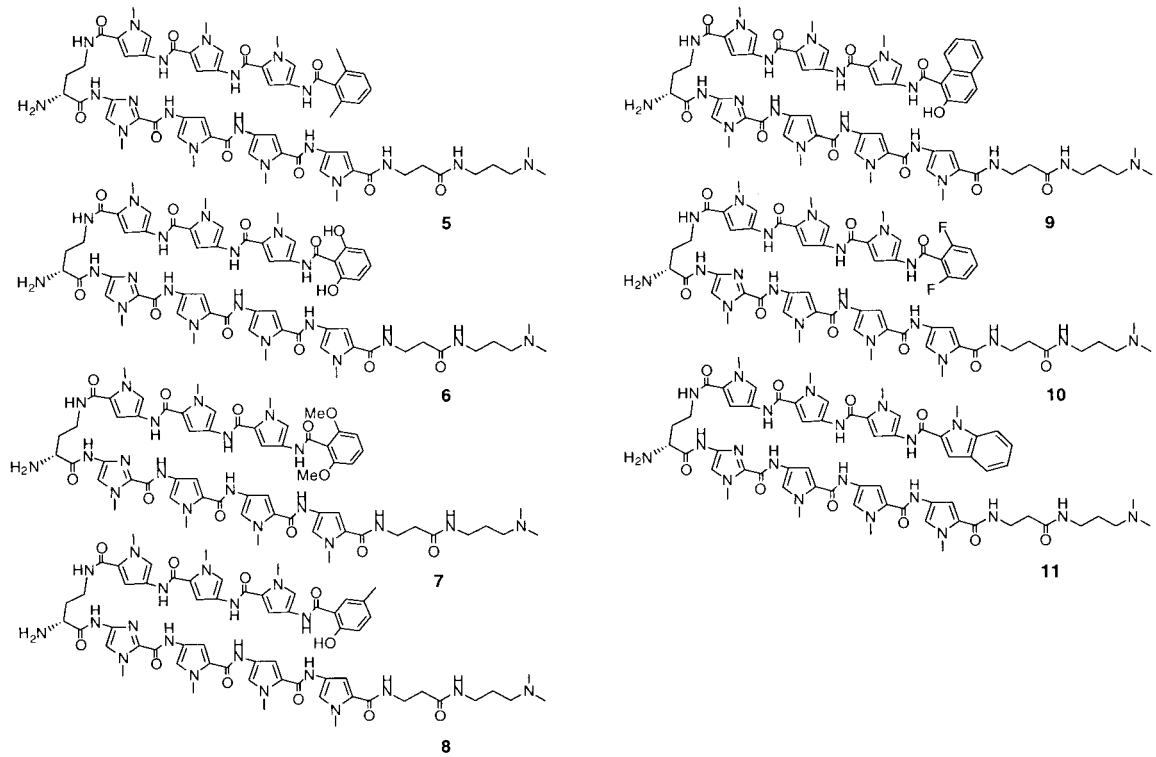


Figure 3.2b. Structures of polyamides **5**-**11**.

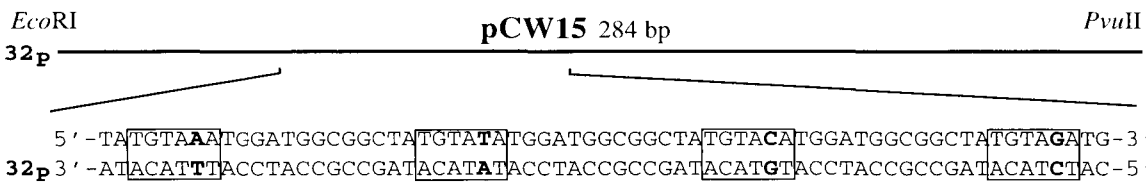


Figure 3.3 284 base pair *Eco*RI/*Pvu*II restriction fragment derived from plasmid pCW15. The targeted six-base pair recognition sites are shown in boxes.

Results

Polyamide synthesis. Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin was synthesized in a stepwise manner from Boc-β-alanine-PAM resin (0.55 mmol/g) using solid-phase methodology[5] in 16 steps (Figure 3.5). Sterically hindered acids such as Hb-**2** and Hb-**3** were coupled as acid chlorides. A sample of the resin was then cleaved by aminolysis with ((dimethyl)amino)propylamine (55°C, 16 h) and purified by reversed-phase HPLC to provide BzPyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**1**). Methyl-protected polyamides **2-4** were subsequently deprotected by treatment with sodium ethanethiolate in DMF (100°C, 1 h)[3] and purified by reversed-phase HPLC to afford (Hb-**1**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**2**), (Hb-**2**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**3**) and (Hb-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**4**) (Figure 3.5). A deprotection study of the dimethoxybenzene compound (**7**) showed a maximum yield of the monomethoxy compound (**4**) after 30 minutes. (Figure 3.4)

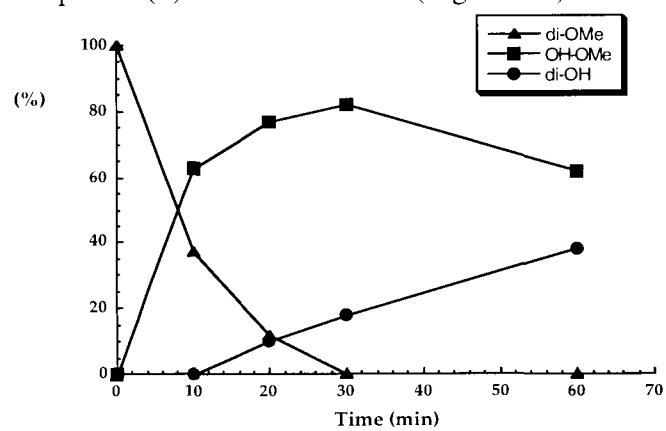


Figure 3.4. Deprotection study of compound **7**. The dimethoxy compound **7** was deprotected at 100°C with EtS^- and samples were taken out at different times and analyzed by HPLC. The different amounts of the dimethoxy-, monomethoxy- and the diol were measured.

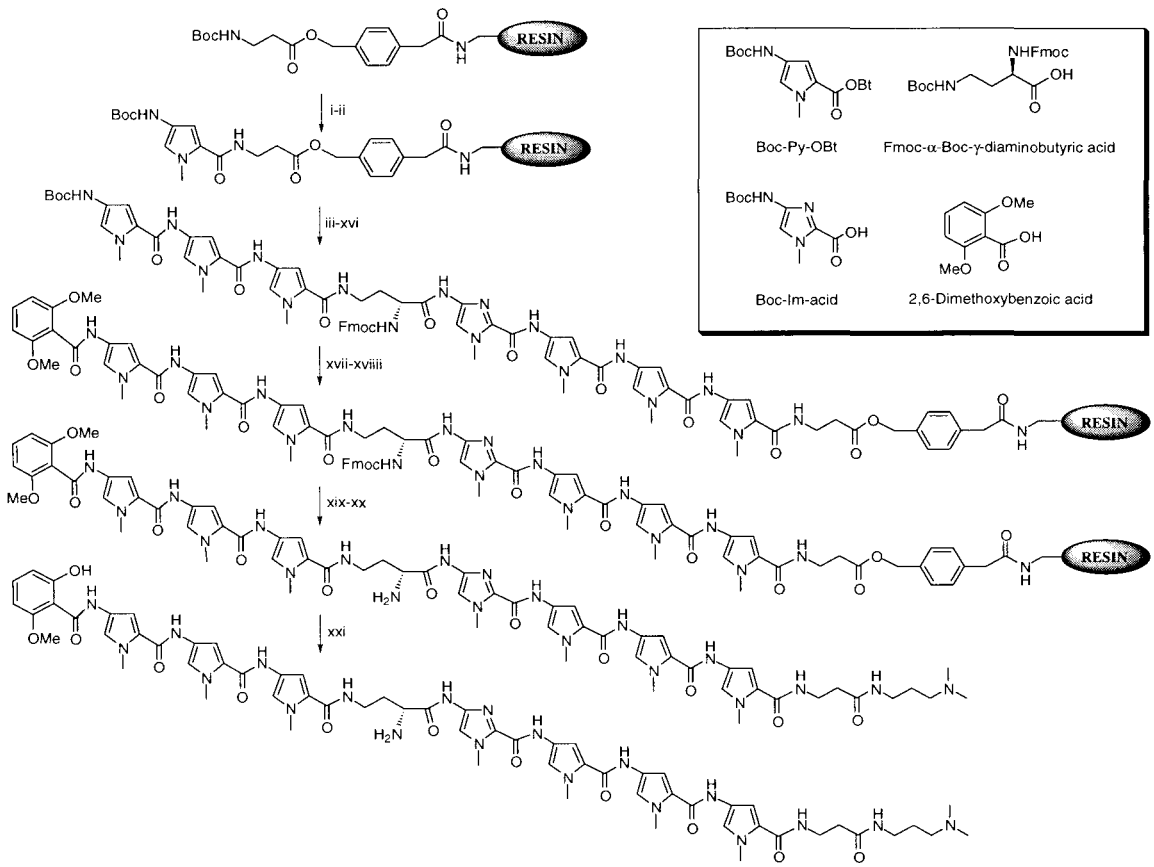


Figure 3.5. Solid-phase synthetic scheme for (Hb-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**4**) starting from commercially available Boc-β-PAM-resin: (i) 80% TFA/DCM, 0.4 M PhSH. (ii) Boc-Py-OBt, DIEA, DMF. (iii) 80% TFA/DCM, 0.4 M PhSH. (iv) Boc-Py-OBt, DIEA, DMF. (v) 80% TFA/DCM, 0.4 M PhSH. (vi) Boc-Py-OBt, DIEA, DMF. (vii) 80% TFA/DCM, 0.4 M PhSH. (viii) Boc-Im-acid, DCC, HOBt, DIEA, DMF. (ix) 80% TFA/DCM, 0.4 M PhSH. (x) Fmoc-α-Boc-γ-diaminobutyric acid, HBTU, 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-Py-OBt, DIEA, DMF. (xvii) 80% TFA/DCM, 0.4 M PhSH. (xviii) 2,6-Dimethoxybenzoic acid, DMF, DCM, oxalylchloride, DIEA, r.t. 3h. (xix) Piperidine:DMF 3:1. (xx) (Dimethylamino)propylamine, 55°C, 16h.(xxi) NaH, EtSH, DMF, 100°C, 1h.

Quantitative DNase I footprinting titrations. Quantitative DNase I footprint titrations on the 3'-^{32}P-end-labeled 284 bp pCW15 *Eco*RI/*Pvu*II restriction fragment (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$, pH 7.0, 22°C) (Figure 3.6) reveals that BzPyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (**1**) binds the 5'-T**T**TACA-3'and 5'-T**A**TACA-3'sites with comparable equilibration association constants, $K_a = 1.9\times 10^{11}$ M^{-1} and $K_a = 1.2 \times 10^{11}$ M^{-1} respectively. Polyamide **1** binds the two sites 5'-T**G**TACA-3' and 5'-T**C**TACA-3' with five and ten-fold lower affinity, respectively, revealing that the Bz/Py pair is specific for both T•A and A•T over G•C and C•G.

Three polyamides containing 2-hydroxybenzamides at the N-terminal position (Hb-**1**)PyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (**2**), (Hb-**2**)PyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (**3**), (Hb-**3**)PyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (**4**) were tested for their ability to discriminate 5'-T**T**TACA-3' vs 5'-T**A**TACA-3' (Table 3.1). (Hb-**1**)PyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (**2**), containing a single hydroxyl bound both 5'-T**T**TACA-3' and 5'-T**A**TACA-3' with the same affinity ($K_a = 1.9 \times 10^{10}$ M^{-1}). (Hb-**2**)PyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (**3**), with a 2-hydroxyl and a 6-methyl substitution also bound the two sites with equal affinity. Discrimination of 5'-T**T**TACA-3' from 5'-T**A**TACA-3' was achieved by replacing the 6-methyl group in polyamide **5** with 6-methoxy. (Hb-**3**)PyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (**4**) bound 5'-TTTACA-3' with an affinity of Ka = 1.9×10^{10} M-1 and 5'-TATACA-3' with three-fold lower affinity. Importantly, polyamide **4** bound to 5'-TGTACA-3' and to 5'-TCTACA-3' with 55 and 200 times lower affinity, respectively.

Equilibration association constants for compounds **5-11** are given in Table 3.2.

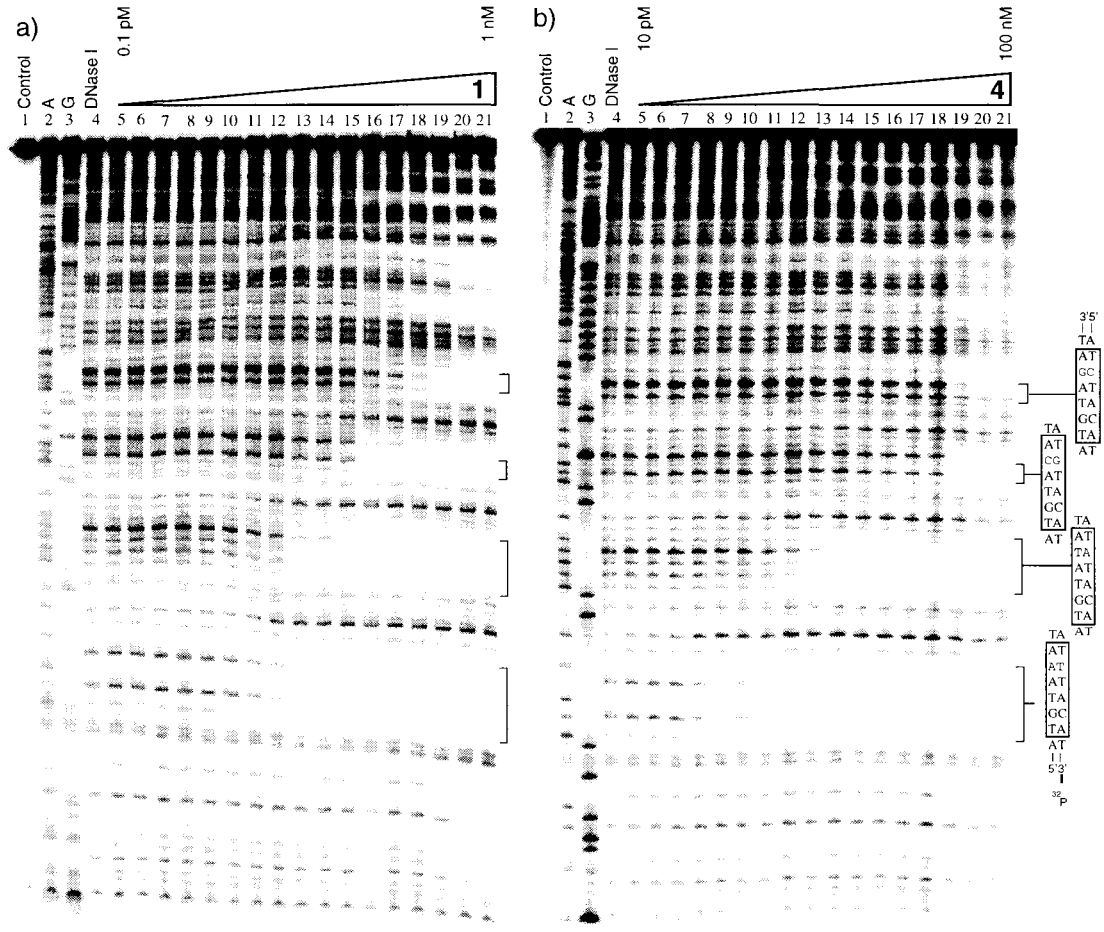


Figure 3.6. (a) Quantitative DNase I footprint titration experiment with BzPyPyPy-(R)$^{H}_{2}N$γ-ImPyPyPy-β-Dp (**1**) on the 284 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW15: lane 1, intact DNA; lane 2, A specific reaction; lane 3, G specific reaction lane 4, DNase I standard; lanes 5-21, 0.1 pM, 0.2 pM, 0.5 pM, 1 pM, 2 pM, 5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 100 pM, 150 pM, 250 pM, 450 pM, 650 pM, 1 nM BzPyPyPy-(R)$^{H}_{2}N$γ-ImPyPyPy-β-Dp (**1**); The four sites 5'-TTTACA-3', 5'-TATACA-3', 5'-TGTACA-3' and 5'-TCTACA-3' sites that were analyzed are shown on the right side of the gel. (b) Quantitative DNase I footprint titration experiment with (Hb-**3**)PyPyPy-(R)$^{H}_{2}N$γ-ImPyPyPy-β-Dp (**4**) on the 284 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW15: lane 1, intact DNA; lane 2, A specific reaction lane 3, G specific reaction lane 4, DNase I standard; lanes 5-21, 10 pM, 20 pM, 50 pM, 100 pM, 150 pM, 250 pM, 400 pM, 650 pM, 1 nM, 1.5 nM, 2.5 nM, 4 nM, 6.5 nM, 10 nM, 20 nM, 50 nM, 100 nM (Hb-**3**)PyPyPy-(R)$^{H}_{2}N$γ-ImPyPyPy-β-Dp (**7**); The four sites 5'-TTTACA-3', 5'-TATACA-3', 5'-TGTACA-3' and 5'-TCTACA-3' sites that were analyzed are shown on the right side of the gel.

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

Polyamide	5'-TTTACA-3'	5'-TATACA-3'	5'-TGTACA-3'	5'-TCTACA-3'
1	$1.9\ (\pm0.3) \times 10^{11}$	$1.2\ (\pm0.1) \times 10^{11}$	$3.5\ (\pm1.0) \times 10^{10}$	$1.7\ (\pm0.7) \times 10^{10}$
2	$1.9\ (\pm0.1) \times 10^{10}$	$1.9\ (\pm0.3) \times 10^{10}$	$6.3\ (\pm1.7) \times 10^{9}$	$3.7\ (\pm0.6) \times 10^{9}$
3	$1.0\ (\pm0.2) \times 10^{10}$	$1.0\ (\pm0.1) \times 10^{10}$	$1.5\ (\pm0.6) \times 10^{9}$	$\leq 1.0 \times 10^{8}$
4	$2.6\ (\pm0.9) \times 10^{10}$	$8.4\ (\pm1.0) \times 10^{9}$	$4.7\ (\pm0.5) \times 10^{8}$	$\leq 1.0 \times 10^{8}$

[a]Values reported are mean values from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. The assays were perfomed at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

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

Polyamide	5'-TTTACA-3'	5'-TATACA-3'	5'-TGTACA-3'	5'-TCTACA-3'
5	1.5×10^{9}	5.7×10^{8}	5.5×10^{8}	1.3×10^{8}
6	2.1×10^{10}	3.9×10^{10}	5.4×10^{9}	1.6×10^{9}
7	4.2×10^{9}	1.5×10^{9}	4.9×10^{8}	2.1×10^{8}
8	4.6×10^{10}	4.0×10^{10}	1.1×10^{10}	8.3×10^{9}
9	6.1×10^{9}	7.0×10^{9}	1.1×10^{9}	1.0×10^{9}
10	2.6×10^{11}	1.1×10^{11}	4.0×10^{10}	3.5×10^{10}
11	4.8×10^{10}	5.7×10^{10}	2.8×10^{10}	6.1×10^{9}

[a]The assays were performed at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

Discussion

BzPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**1**) demonstrates that a simple benzene ring is able to pair with a 1-methylpyrrole ring (Bz/Py) to distinguish A•T/T•A base pairs from G•C/C•G base pairs in the minor groove of DNA. The Bz/Py pair is a higher affinity binder than the Py/Py pair at the terminal position. Interestingly, the Hb-**1**/Py pair with a single hydroxyl group did not discriminate T from A. The hydroxyl group is likely in the wrong conformation, oriented away from the floor of the minor groove and forming a hydrogen bond with the proximal carbonyl group. In order to force the hydroxyl group toward the minor groove of the DNA by an unfavorable steric interaction, two 2-hydroxybenzamides substituted at the 6 position with methyl- and methoxy- were examined. Quantitative footprint titration reveal that T versus A discrimination can be

achieved with the 6-methoxy but not 6-methyl derivative. These results suggest that the 6-methyl substituent projecting to the floor of the helix is not large enough to prevent the hydroxyl group from hydrogen bonding to the carbonyl group. Presumably the steric bulk of the 6-methoxy group favors the rotamer with the hydroxyl group oriented toward the minor groove (Figure 3.7). Polyamide **4** with the Hb-**3**/Py pair binds 5'-TTTACA-3' and 5'-TATACA-3' with binding affinities of $K_a = 2.6 \times 10^{10}\ M^{-1}$ and $K_a = 8.4 \times 10^{9}\ M^{-1}$, respectively. Although the three-fold T versus A selectivity is similar to the Hp-**3**/Py pair, the overall affinity is significantly higher, and most importantly, the selectivity of A/T vs G/C is dramatically improved. The new hydroxybenzamide/pyrrole pair (Hb/Py) is a step towards hairpin polyamides capable of differentiating pure A•T tract sequences important in gene regulation studies.

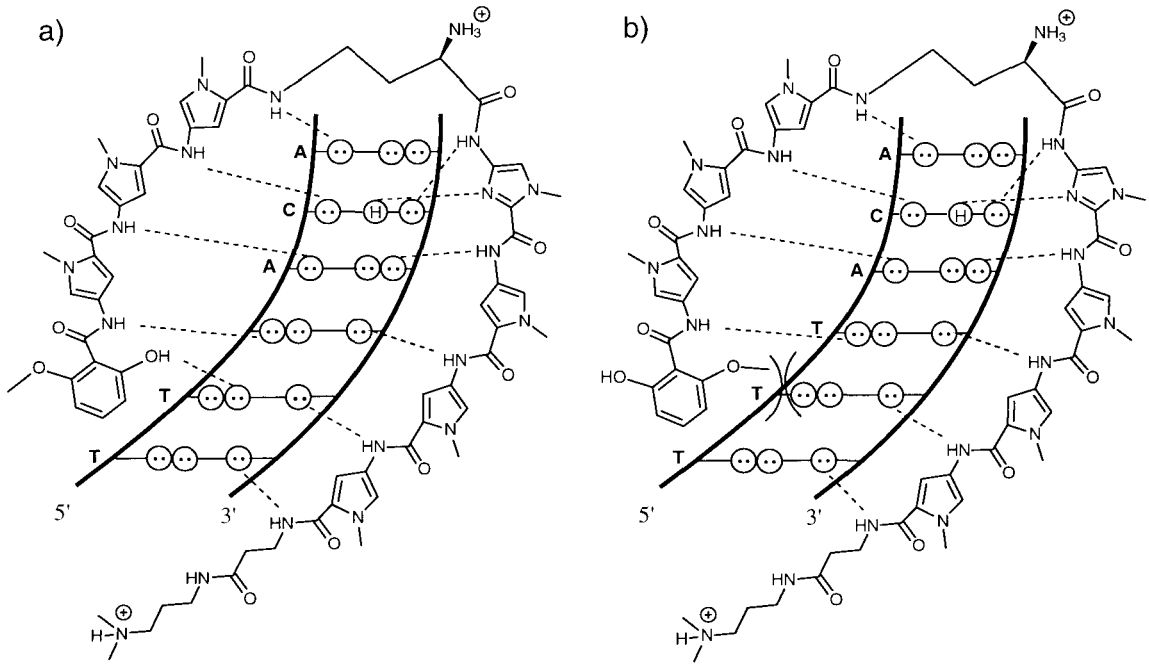


Figure 3.7. Binding models for two rotamers of (Hb-**3**)PyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp and a 5'-TTTACA-3' sequence.

Experimental Section

Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-PAM-Resin), N,N'-Dicyclohexylcarbodiimide (DCC), N-Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and Boc-β-alanine were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA) and *N,N*-dimethylformamide (DMF), were purchased from Applied Biosystems. The DMF was distilled under reduced pressure prior to synthesis. (R)-2-Fmoc-4-Boc-diaminobutyric acid was from Bachem, dichloromethane (DCM) was reagent grade from EM, thiophenol (PhSH) and dimethylaminopropylamine from Aldrich, and trifluoroacetic acid (TFA) from Halocarbon. ^{1}H NMR were recorded on a Varian Mercury 300 instrument. Chemical shifts are reported in ppm downfield from the signal for Me_4Si, with reference 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 on a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5μm, 300 × 4.6 mm reversed-phase column in 0.1% (w/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 × 100 mm, 100μm C_{18} column, 0.1% (w/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.

Boc-PyPyPy-(R)$^{Fmoc}\gamma$-ImPyPyPy-β-PAM-resin. Boc-PyPyPy-(R)$^{Fmoc}\gamma$-ImPyPyPy-β-PAM-resin (0.335 mmol/g) was synthesized, as previously reported[3], in a stepwise fashion from 0.55 mmol/g Boc-β-PAM-resin by manual solid-phase methods.

Procedure for cleavage from the resin. After the coupling was completed, the resin was filtered off the reaction and washed with DMF (2 × 30 seconds). DMF (1 mL) and piperidine (3 mL) were added and the mixture shaken for 30 minutes at 22°C. The resin was filtered off and washed with DMF (2 × 30 seconds), DCM (3 × 30 seconds), MeOH (1 × 30 seconds), Et_2O (1 × 30 seconds) and dried *in vacuo*. The resin was treated with (dimethylamino)propylamine (1 mL) with periodic agitation at 55°C for 16 hours. The reaction mixture was then filtered to remove the resin, TFA (0.1% (w/v), 7 mL) added and the resulting solution purified by reversed-phase HPLC. The pure compounds were recovered as white powders upon lyophilization of the appropriate fractions.

Deprotection of methyl-protected polyamides. To a slurry of NaH (80 mg, 60% in mineral oil) in DMF (0.50 mL) was added a solution of ethanethiol (0.32 mL) in DMF (0.50 mL) and the mixture was heated to 100°C for 5 minutes in a sealed tube. The polyamide, dissolved in DMF (1.00 mL), was added to the ethanethiolate solution and the mixture was heated to 100°C for 1 hour in a sealed tube, and then cooled to 0°C. The mixture was added to HOAc (3 mL) and all volatiles were removed (high vacuum, 40°C). The residue was dissolved in CH_3CN (1 mL) and TFA (7 mL, 0.1% (w/v)) and purified by preparative HPLC.

BzPyPyPy-(R)$^{H_2N}\gamma$-ImPyPyPy-β-Dp (1). Benzoic acid (70 mg, 0.57 mmol) was dissolved in DMF (1.0 mL) and HBTU (75 mg, 0.20 mmol) and DIEA (0.5 mL) were added. The mixture was allowed to stir for 10 minutes and added to the deprotected Boc-PyPyPy-(R)$^{Fmoc}\gamma$-ImPyPyPy-β-PAM-resin (100 mg). The mixture was shaken for 5.0 hours at 22°C, and the polyamide was cleaved from the resin according to the general procedure. Preparative HPLC gave **1** (16 mg, 39% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^1H

NMR (DMSO-d_6) δ 11.10, 10.42, 10.20, 10.07, 10.04, 10.02, 9.97 (s, 1H each), 9.41 (bs, 1 H), 8.42 (bs, 2 H), 8.26 (t, 1 H, J = 5.7 Hz), 8.12 (q, 2 H, J = 5.6 Hz), 8.00 (dd, 2 H, J = 6.3, 1.5 Hz), 7.56-7.70 (m, 3 H), 7.40 (d, 1 H, J = 1.5 Hz), 7.33 (d, 1 H, J = 1.5 Hz), 7.30 (d, 1 H, J = 1.8 Hz), 7.28 (d, 1 H, J = 1.5 Hz), 7.26 (d, 1 H, J = 2.1 Hz), 7.25 (d, 1 H, J = 2.1 Hz), 7.21 (d, 1 H, J = 2.1 Hz), 7.18 (d, 1 H, J = 1.8 Hz), 7.15 (d, 2 H, J = 1.5 Hz), 7.05 (d, 1 H, J = 2.1 Hz), 6.95 (d, 1 H, J = 1.5 Hz), 4.05, 3.94, 3.93, 3.92, 3.91, 3.88, 3.87 (s, 3 H each), 3.30-3.50 (m, 4 H), 3.17 (q, 2 H, J = 6.3 Hz), 3.06 (p, 2 H, J = 4.8 Hz), 2.81 (d, 6 H, J = 4.8 Hz), 2.42 (t, 2 H, J = 7.2 Hz), 2.00-2.10 (m, 2H), 1.81 (p, 2 H, J = 7.5 Hz); MALDI-TOF-MS calcd. for $C_{60}H_{72}N_{20}O_{10}$ (M+H): 1233.6; found: 1233.7.

(Hb-1)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (2). 2-Methoxybenzoic acid (70 mg, 0.46 mmol) was dissolved in DMF (1.0 mL) and HBTU (75 mg, 0.20 mmol) and DIEA (0.5 mL) were added. The mixture allowed to stir for 10 minutes and then added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (100 mg). The mixture was shaken for 4.5 hours at 22°C and the polyamide was cleaved from the resin according to the general procedure. Preparative HPLC gave methylprotected (Hb-**1**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (15 mg, 35% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 11.10, 10.19, 10.07 (s, 1 H each), 10.03 (s, 2 H), 10.01, 9.97 (s, 1 H each), 8.38-8.50 (bs, 2 H), 8.26 (t, 1 H, J = 6.0 Hz), 8.12 (q, 2 H, J = 6.0 Hz), 7.72 (d, 1 H, J = 1.8 Hz), 7.69 (d, 1 H, J = 1.8 Hz), 7.61 (s, 1 H), 7.58 (d, 1 H, J = 2.1 Hz), 7.55 (bs, 1 H), 7.52 (d, 1 H, J = 1.8 Hz), 7.38 (d, 1 H, J = 1.8 Hz), 7.33 (d, 1 H, J = 1.5 Hz), 7.29 (d, 1 H, J = 1.5 Hz), 7.28 (d, 1 H, J = 1.5 Hz), 7.25 (s, 2 H), 7.22 (s, 1 H), 7.10-7.20 (m, 4 H), 7.05 (d, 1 H, J = 1.5 Hz), 6.95 (d, 1 H, J = 1.5 Hz), 4.05, 3.96, 3.93, 3.92, 3.91,3.90, 3.88, 3.87 (s, 3 H each), 3.30-3.50 (m, 4 H), 3.17 (q, 2 H, J = 6.0 Hz), 3.07 (p, 2 H, J = 5.7 Hz), 2.80 (d, 6 H, J = 4.8 Hz), 2.42 (t, 2 H, J = 7.2 Hz), 2.06 (bs, 2 H), 1.80 (p, 2H, J = 8.7 Hz); MALDI-TOF-MS calcd. for $C_{61}H_{74}N_{20}O_{11}$ (M+H): 1263.6; found: 1263.7. The methylprotected (Hb-**1**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (3.1 mg, 2.5 μmol) was deprotected according to the general procedure to give **2** (0.5 mg, 16% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR

(DMSO-d_6) δ 12.25, 10.96, 10.43, 10.06, 9.92, 9.90, 9.88, 9.83 (s, 1 H each), 8.25 (bs, 3 H), 8.13 (t, 1 H, J = 5.9 Hz), 7.98 (q, 2 H, J = 6.0 Hz), 7.92 (dd, 1 H, J = 8.4, 1.8 Hz), 7.48 (s, 1 H), 7.37 (ddd, 1 H, J = 8.7, 7.2, 1.5 Hz), 7.28 (d, 1 H, J = 1.5 Hz), 7.20 (d, 1 H, J = 1.8 Hz), 7.17 (d, 1 H, J = 1.5 Hz), 7.14 (d, 1 H, J = 1.5 Hz), 7.11 (bs, 2 H), 7.08 (d, 1 H, J = 1.5 Hz), 7.07 (d, 1 H, J = 2.1 Hz), 7.02 (d, 2 H, J = 1.5 Hz), 6.85-6.95 (m, 3 H), 6.82 (d, 1 H, J = 1.5 Hz), 3.92, 3.82, 3.79, 3.78, 3.77, 3.75, 3.73 (s, 3 H each), 3.04 (q, 2 H, J = 6.6 Hz), 2.93 (bt, 2 H, J = 7.2 Hz), 2.68 (bs, 6 H), 2.28 (t, 2 H, J = 7.2 Hz), 1.92 (bq, 2 H, J = 7.7 Hz), 1.67 (p, 2 H, J = 7.8 Hz); MALDI-TOF-MS calcd. for $C_{60}H_{72}N_{20}O_{11}$ (M+H): 1249.6; found: 1249.7.

(Hb-2)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (3). 2-Methoxy-6-methylbenzoic acid (26 mg, 0.16 mmol) was dissolved in DMF/DCM (1:1, 0.2 mL) and added to a mixture of DMF (0.1 mL) and oxalylchloride (2M in DCM, 0.072 mL). The mixture was allowed to stir for 20 minutes and DMF (0.3 mL) and DIEA (0.2 mL) were added. The mixture was then added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (50 mg) and shaken for 4.0 hours at 22°C. The polyamide was cleaved from the resin according to the general procedure. Preparative HPLC afforded methylprotected (Hb-**2**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (9 mg, 41% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 10.27, 10.19 (s, 1 H each), 10.03 (s, 2 H), 10.00, 9.97 (s, 1 H each), 8.26 (t, 1 H, J = 6.0 Hz), 8.12 (q, 2 H, J = 5.1 Hz), 7.61 (s, 1 H), 7.30-7.40 (m, 3 H), 7.29 (d, 1 H, J = 1.5 Hz), 7.28 (d, 1 H, J = 2.4 Hz), 7.25 (bs, 2 H), 7.22 (d, 1 H, J = 1.8 Hz), 7.15 (bs, 2 H), 7.04, 7.03, 7.00, 6.96 (s, 1 H each), 6.95 (d, 1 H, J = 1.5 Hz), 6.93, 6.91 (s, 1 H each), 4.05 (2, 3 H), 3.92, 3.91 (s, 6 H each), 3.88, 3.87, 3.81 (s, 3 H each), 3.30-3.50 (m, 4 H), 3.17 (q, 2 H, J = 6.0 Hz), 3.05 (bt, 2 H, J = 7.6 Hz), 2.79 (s, 6H), 2.42 (t, 2 H, J = 7.2 Hz), 2.27 (s, 3 H), 2.06 (bs, 2H), 1.80 (p, 2 H, J = 8.1 Hz); MALDI-TOF-MS calcd. for $C_{62}H_{76}N_{20}O_{11}$ (M+H): 1277.6; found: 1277.7. The methylprotected (Hb-**1**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (3.1 mg, 2.4 μmol) was deprotected according to the general procedure to give **3** (1.4 mg, 48% recovery). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 10.19, 10.18, 10.03, 10.02,

10.00, 9.96, 9.63 (s, 1 H each), 8.25 (bt, 1 H, J = 5.8 Hz), 8.12 (q, 2 H, J = 3.4 Hz), 7.61 (s, 1 H), 7.32 (d, 1 H, J = 1.7 Hz), 7.31 (d, 1 H, J = 1.8 Hz), 7.29 (d, 1 H, J = 2.1 Hz), 7.28 (d, 1 H, J = 2.1 Hz), 7.24 (d, 2 H, J = 1.8 Hz), 7.22 (d, 1 H, J = 1.5 Hz), 7.10-7.20 (m, 3 H), 7.06 (d, 1 H, J = 2.1 Hz), 7.04 (d, 1 H, J = 1.8 Hz), 6.95 (d, 1 H, J = 1.5 Hz), 6.78 (d, 1 H, J = 8.1 Hz), 6.73 (d, 1 H, J = 7.2 Hz), 4.05 (s, 3 H), 3.92, 3.91 (s, 6 H each), 3.88, 3.87 (s, 3 H each), 3.17 (q, 2 H, J = 6.0 Hz), 3.04 (bs, 2 H), 2.78 (s, 6 H), 2.41 (t, 2 H, J = 6.9 Hz), 2.25 (s, 3 H), 2.05 (bs, 2 H), 1.79 (p, 2 H, J = 6.6 Hz); MALDI-TOF-MS calcd. for $C_{61}H_{74}N_{20}O_{11}$ (M+H): 1263.6; found: 1263.7.

(Hb-3)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (4). 2,6-Dimethoxybenzoic acid (28 mg, 0.16 mmol) was dissolved in DMF/DCM (1:1, 0.2 mL) and added to a mixture of DMF (0.1 mL) and oxalylchloride (2M in DCM, 0.072 mL). The mixture was allowed to stir for 20 minutes and DMF (0.3 mL) and DIEA (0.2 mL) were added. The mixture was added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (50 mg) and shaken for 3.0 hours at 22°C. The polyamide was then cleaved from the resin according to the general procedure. Preparative HPLC gave methylprotected (Hb-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (10 mg, 48% recovery). UV λmax (H_2O) 240, 308 (69500); 1H NMR (DMSO-d6) δ 10.96, 10.06, 10.03, 9.90, 9.88, 9.87, 9.83 (s, 1 H each), 8.26 (bs, 3 H), 8.12 (t, 1 H, J = 6.2 Hz), 7.99 (q, 2 H, J = 5.7 Hz), 7.48 (s, 1 H), 7.27 (t, 2 H, J = 8.4 Hz), 7.20 (d, 1 H, J = 1.6 Hz), 7.16 (d, 1 H, J = 1.8 Hz), 7.14 (d, 1 H, J = 2.1 Hz), 7.13 (d, 1 H; J = 1.8 Hz), 7.11 (d, 1 H, J = 1.5 Hz), 7.08 (d, 1 H, J = 1.5 Hz), 7.02 (d, 1 H, J = 1.9 Hz), 7.01 (d, 1 H, J = 1.6 Hz), 6.90 (d, 1 H, J = 1.5 Hz), 6.89 (d, 1 H, J = 2.1 Hz), 6.82 (d, 1 H, J = 1.8 Hz), 6.66, 6.63 (s, 1 H each), 3.92, 3.79, 3.78 (s, 3 H each), 3.78 (s, 6 H), 3.75, 3.74 (s, 3 H each), 3.67 (s, 6 H), 3.04 (q, 2 H, J = 6.0 Hz), 2.94 (p, 2 H, J = 4.8 Hz), 2.68 (d, 6 H, J = 4.5 Hz), 2.28 (t, 2 H, J = 6.9 Hz), 1.94 (bs, 2 H), 1.67 (p, 2 H, J = 7.6 Hz); MALDI-TOF-MS calcd. for $C_{62}H_{76}N_{20}O_{12}$ (M+H): 1293.6; found: 1293.7. The methylprotected (Hb-**3**)PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (2.7 mg, 2.1 μmol) was deprotected according to the general procedure to give **4** (1.9 mg, 42% recovery) together with (2,6-dihydroxybenzoyl)HN-

PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (1.3 mg, 28% recovery). UV λ_{max} (H_2O) 240, 310 (69500); 1H NMR (DMSO-d_6) δ 11.84, 10.08, 10.06, 9.89, 9.88, 9.83 (s, 1 H each), 8.25 (bs, 3 H), 8.12 (t, 1 H, J = 5.7 Hz), 7.98 (q, 2 H, J = 4.8 Hz), 7.48 (s, 1 H), 7.18-7.28 (m, 3 H), 7.16 (d, 1 H, J = 1.5 Hz), 7.14 (d, 1 H, J = 1.5 Hz), 7.11 (bs, 2 H), 7.08 (d, 1 H, J = 1.5 Hz), 7.02 (d, 2 H, J = 1.5 Hz), 7.01 (d, 1H, J = 1.8 Hz), 6.91 (d, 1 H, J = 1.5 Hz), 6.83 (d, 1 H, J = 1.8 Hz), 6.52 (d, 1H, J = 7.8 Hz), 6.47 (d, 1 H, J = 8.4 Hz), 3.92 (s, 3 H), 3.80 (s, 6 H), 3.79, 3.78, 3.77, 3.75, 3.73 (s, 3H each), 3.04 (q, 2 H, J = 6.0 Hz), 2.94 (p, 2 H, J = 4.8 Hz), 2.68 (d, 6 H, J = 4.2 Hz), 2.28 (t, 2 H, J = 7.5 Hz), 1.94 (bq, 2H, J = 8.4 Hz), 1.67 (p, 2 H, J = 8.7 Hz); MALDI-TOF-MS calcd. for $C_{61}H_{74}N_{20}O_{12}$ (M+H): 1279.6; found: 1279.7.

2,6-DimethylbensoylHN-PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (5).

2,6-Dimethylbenzoic acid (24 mg, 0.16 mmol) was dissolved in DMF/DCM (1:1, 0.2 mL) and added to a mixture of DMF (0.1 mL) and oxalylchloride (2M in DCM, 0.072 mL). The mixture was stirred for 20 minutes and DMF (0.3 mL) and DIEA (0.2 mL) were added. The mixture was then added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (50 mg) and shaken for 4.0 hours at 22°C. Preparative HPLC gave **5** (6 mg, 27% recovery). UV λ_{max} (H_2O) 240, 312 (69500); MALDI-TOF-MS calcd. for $C_{62}H_{76}N_{20}O_{10}$ (M+H): 1261.6; found: 1261.8.

2,6-Dihydroxy-bensoylHN-PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (6).

Methyl-protected **4** (2.1 mg, 1.6 μmol) was deprotected according to the general procedure but heated for 7.5 hours. Preparative HPLC gave **6** (0.6 mg, 29% recovery). UV λ_{max} (H_2O) 240, 310 (69500); MALDI-TOF-MS calcd. for $C_{60}H_{72}N_{20}O_{12}$ (M+H): 1265.6; found: 1265.5.

(2-Hydroxy-1-naphtoyl)HN-PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (9).

2-Methoxy-1-naphtoic acid[6] (32 mg, 0.16 mmol) was dissolved in DMF/DCM (1:1, 0.2 mL) and added to a mixture of DMF (0.1 mL) and oxalylchloride (2M in DCM, 0.072 mL). The mixture was stirred for 20 minutes and DMF (0.3 mL) and DIEA (0.2 mL) were added. The mixture was then added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin

(50 mg) and shaken for 3.0 hours at 22°C. Preparative HPLC gave methylprotected **9** (7 mg, 33% recovery). UV λ_{max} (H_2O) 230, 314 (69500); ^1H NMR (DMSO-d_6) δ ; MALDI-TOF-MS calcd. for $C_{65}H_{76}N_{20}O_{11}$ (M+H): 1313.6; found: 1313.7. The methylprotected compound (3.4 mg, 2.6 μmol) was treated according to the general procedure. Preparative HPLC gave **9** (1.8 mg, 54% recovery). UV λ_{max} (H_2O) 230, 314 (69500) ; MALDI-TOF-MS calcd. for $C_{64}H_{74}N_{20}O_{11}$ (M+H): 1299.6; found: 1299.7.

2,6-Difluoro-bensoylHN-PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (10).

2,6-Difluoro-benzoic acid (70 mg, 0.44 mmol) was dissolved in DMF (1.0 mL) and HBTU (75 mg, 0.20 mmol) and DIEA (0.5 mL) were added and the mixture stirred for 10 minutes and added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (100 mg) and the mixture shaken for 5.0 hours at 22°C. Preparative HPLC gave **10** (8 mg, 20% recovery). UV λ_{max} (H_2O) 240, 307 (69500); MALDI-TOF-MS calcd. for $C_{60}H_{70}F_2N_{20}O_{10}$ (M+H): 1269.6; found: 1269.7.

Indolyl-PyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (11).

1-Methyl indole-2-carboxylic acid (57 mg, 0.32 mmol) was dissolved in DMF (1.0 mL) and HOBt (30 mg, 0.22 mmol) and DCC (46 mg, 0.23) were added and the mixture stirred for 10 minutes and added to the deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (100 mg) followed by DIEA (0.5 mL) and the mixture shaken for 3.0 hours at 22°C. Preparative HPLC gave **11** (10 mg, 24% recovery). UV λ_{max} (H_2O) 240, 312 (69500); MALDI-TOF-MS calcd. for $C_{63}H_{75}N_{21}O_{10}$ (M+H): 1286.6; found: 1286.6.

Quantitative DNase I Footprinting. DNase I Footprinting reactions were carried out on the 284 bp *Eco*RI / *Pvu*II restriction fragment of pCW15 as previously described.[3]

Acknowledgement

We are grateful to the National Institutes of Health (GM-27681) for research support and a research traineeship award (GM-08501) to C.C.C.W. and the Knut and Alice

Wallenberg Foundation for a post-doctoral fellowship to U.E. We thank G. M. Hathaway for MALDI-TOF mass spectrometry.

References

1. Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688-693.
2. (a) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111-115. (b) Kielkopf, C. L.; Bremer, R. B.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557-567.
3. Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 9354-9360.
4. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.
5. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
6. Ahmed, A.; Bragg, R. A.; Clayden, J.; Lai, L. W.; McCarthy, C.; Pink, J. H.; Westlund, N.; Yasin, S. A. *Tetrahedron*, **1998**, *54*, 13277-13294.

Chapter 4

Expanding the Recognition of the Minor Groove of DNA by Hairpin Polyamides to Include 5'-CWWC-3' Sequences

The text of this chapter was taken in part from a publication coauthored with Dr. Ulf Ellervik and Prof. Peter B. Dervan.
(Wang, C. C. C.; Ellervik, U.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, 3, 653-657.)

Abstract

In order to evaluate hairpin polyamides capable of distinguishing C•G base pairs at the terminal position three polyamides, PyPyPyPy-(R)^{H_2N}γ-ImPyPyPyIm-β-Dp (**1**), PyPyPyPy-(R)^{H_2N}γ-ImPy-β–Im-β-Dp (**2**) and BzPyPyPy-(R)^{H_2N}γ-ImPy-β–Im-β-Dp (**4**) were synthesized. The equilibrium association constants (K_a) were determined at four DNA sites which differ at a single common position, 5'-T**N**TACA-3' (N = T, A, G, C). Quantitative DNase I footprint titration experiments reveal that PyPyPyPy-(R)^{H_2N}γ-ImPyPyPyIm-β-Dp (**1**) binds the four binding sites with similar affinities, K_a = 1.3-1.9 × 10^{10} M^{-1} indicating that there is no preference for any specific base pair. In contrary PyPyPyPy-(R)^{H_2N}γ-ImPy-β–Im-β-Dp (**2**) with a flexible aliphatic β-alanine next to the key imidazole group bound to the match site 5'-T**C**TACA-3' with $K_a = 4.9 \times 10^{10}$ M^{-1} and the single base pair mismatch sites with 5-25-fold lower affinity. Compound **4** showed similar results to compound **2**. These results expand the repertoire of sequences targetable by hairpin to include sequences in promoters beginning with C•G base pairs and will be an important step forward in our gene regulation studies.

Introduction

Polyamides containing three aromatic amino acids, N-methylpyrrole (Py), N-methylimidazole (Im) and 3-hydroxy-1-methylpyrrole (Hp), bind as antiparallel pairs in the minor groove of DNA with affinities and specificities for predetermined sequences comparable to DNA-binding proteins.[1] Im/Py pair binds G•C, while a Py/Im pair recognizes C•G. A Py/Py pair binds both A•T and T•A base pairs in preference to G•C and C•G. An Hp/Py pair prefers T•A over A•T, C•G or G•C.

Polyamide dimers linked by a γ turn unit creates a hairpin structure which sets the ring pairs unambiguously in register.[2] We have observed that hairpin polyamides with Im at the N-terminal position usually have higher affinities than the corresponding Py analogs.[1] The placement of an Im and Py in the first and eighth position of an eight-ring polyamide create an Im/Py pair at the N- and C-terminal positions of the hairpin, biasing our choice of target DNA sequences beginning with a G•C base pair.[1]

We would like to explore ring pairings that would allow the DNA sequence repertoire to be broadened wherein the target sequence begins in a C•G base pair. We have shown that Py/Im pairs when placed *internal* in a hairpin distinguish C•G from G•C.[3] However, this has not been demonstrated for the terminal position in hairpin polyamides where the Im residue would necessarily be placed at the C-terminus. Replacement of a Py ring with a flexible β-alanine amino acid has been demonstrated to be necessary to target GWG sequences.[4] It remains to be determined whether a flexible β-alanine linker would also be necessary to allow targeting of C•G in the terminal position. The ability to target DNA sequences in promoters beginning with C•G base pairs will be an important step forward in our gene regulation studies.[1]

We describe here three eight-ring hairpin polyamides with Py/Im at the N-terminal and C-terminal positions (Figure 4.1). PyPyPyPy-(R)H_2Nγ-ImPyPyIm-β-Dp (**1**) contains eight ring residues and PyPyPyPy-(R)H_2Nγ-ImPy-β-Im-β-Dp (**2**) and BzPyPyPy-

(R)H_2Nγ-ImPy-β-Im-β-Dp (**4**) contain a flexible β-alanine amino acid and seven other rings residues. In addition, ImPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**3**)[5] was included to establish a baseline with previously published work (Figure 4.2). All four eight-ring hairpin polyamides were synthesized by solid-phase methods.[6] The plasmid pCW15[5] was designed to contain the four six-base pair recognition sites 5'-TTTACA-3', 5'-TATACA-3', 5'-TGTACA-3' and 5'-TCTACA-3' which differ at a single common position allowing for comparison of the affinities between different terminal pairs and the four Watson Crick base pairs in the minor groove of DNA (Figure 4.3).

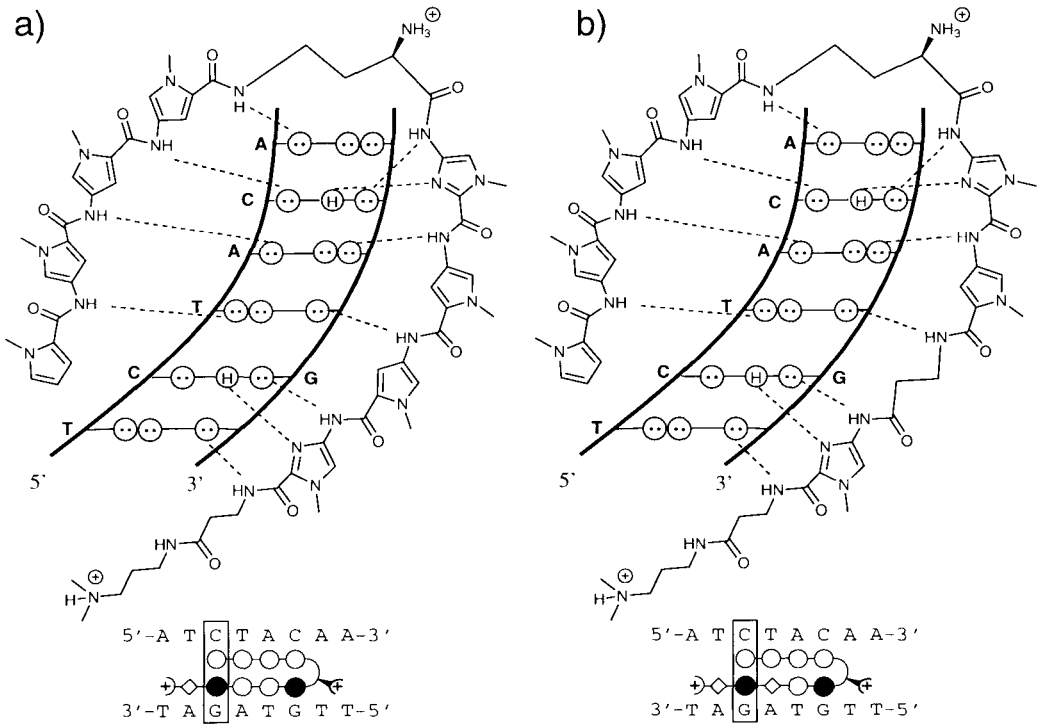


Figure 4.1. (a) Hydrogen bonding model of the polyamide:DNA complex between eight-ring hairpin polyamide PyPyPyPy-(R)H_2Nγ-ImPyPyIm-β-Dp and a 5'-TCTACA-3' site. A circle with two dots represents the lone pairs of N3 of purines and O2 of cytosine. Two touching circles with dots represent the two lone pairs of the O2 of thymine. Circles containing an H represent the exocyclic amino hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (b) Hydrogen bonding model of the polyamide:DNA complex between seven-ring hairpin polyamide PyPyPyPy-(R)H_2Nγ-ImPy-β–Im-β-Dp and a 5'-TCTACA-3' site.

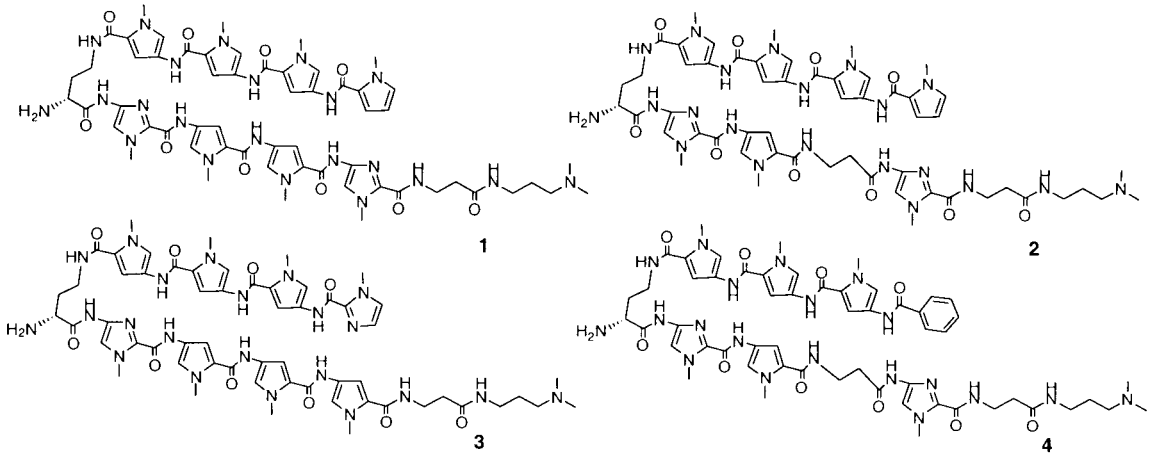


Figure 4.2. Structures of polyamides. PyPyPyPy-(R)H_2Nγ-ImPyPyIm-β-Dp (**1**), PyPyPyPy-(R)H_2Nγ-ImPy-β–Im-β-Dp (**2**), ImPyPyPy-(R)H_2Nγ-ImPyPyPy -β-Dp (**3**), BzPyPyPy-(R)H_2Nγ-ImPy-β–Im-β-Dp (**4**).

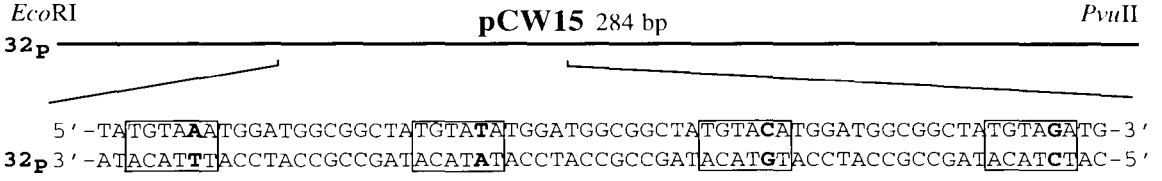


Figure 4.3. 284 base pair *Eco*RI/*Pvu*II restriction fragment derived from plasmid pCW15. The targeted six-base pair recognition sites are shown in boxes.

Results

Polyamide synthesis. PyPyPyPy-(R)^{Fmoc}γ-ImPyPyIm-β-PAM-resin and PyPyPyPy-(R)^{Fmoc}γ-ImPy-β–Im-β-PAM-resin were synthesized in a stepwise manner from Boc-β-alanine-PAM resin (0.55 mmol/g) using solid-phase methodology[6] in 14 steps (Figure 4.4). A sample of the resin was then cleaved by aminolysis with ((dimethyl)amino)propylamine (55°C, 16 hours) and purified by reversed-phase HPLC to provide PyPyPyPy-(R)H_2Nγ-ImPyPyIm-β-Dp (**1**) PyPyPyPy-(R)H_2Nγ-ImPy-β–Im-β-Dp (**2**) and BzPyPyPy-(R)H_2Nγ-ImPy-β-Im-β-Dp (**4**) (Figure 4.2).

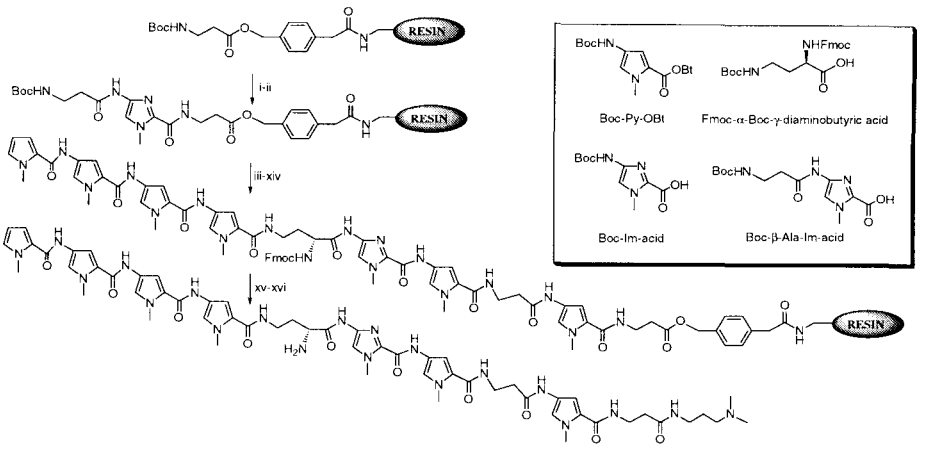


Figure 4.4. Solid-phase synthetic scheme for PyPyPyPy-(R)^{H_2N}γ-ImPy-β–Im-β-Dp (**2**) starting from commercially available Boc-β-PAM-resin: (i) 80% TFA/DCM, 0.4 M PhSH. (ii) Boc-β-Ala-Im-acid, HBTU, DIEA, DMF. (iii) 80% TFA/DCM, 0.4 M PhSH. (iv) Boc-Py-OBt, DIEA, DMF. (iii) 80% TFA/DCM, 0.4 M PhSH. (iv) Boc-Im-acid, DCC, HOBt, DIEA, DMF. (v) 80% TFA/DCM, 0.4 M PhSH. (vi) Fmoc-α-Boc-γ-diaminobutyric acid, HBTU, DIEA, DMF.(vii) 80% TFA/DCM, 0.4 M PhSH. (viii) Boc-Py-OBt, DIEA, DMF. (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) Py-acid, HBTU, DIEA, DMF. (xv) Piperidine:DMF 3:1. (xvi) (Dimethylamino)propylamine, 55°C, 16h.

Quantitative DNase I footprinting titrations. Quantitative DNase I footprint titrations on the 3'-^{32}P-end-labeled 284 bp pCW15 *Eco*RI/*Pvu*II restriction fragment (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$, pH 7.0, 22°C) (Figure 4.5) reveal that PyPyPyPy-(R)^{H_2N}γ-ImPyPyIm-β-Dp (**1**) binds the four binding sites with similar affinities, K_a = 1.3-1.9 × 10^{10} M^{-1}, indicating that there is no preference for any specific base pair (Table 1). In contrary compound **2** with a flexible aliphatic β-alanine next to the key imidazole group bound the 5'-T**C**TACA-3' site with $K_a = 4.9 \times 10^{10}$ M^{-1} and the single base pair mismatched sites 5'-T**G**TACA-3', 5'-T**T**TACA-3'and 5'-T**A**TACA-3' with lower affinities ($K_a = 1.0 \times 10^{10}$ M^{-1}, $K_a = 8.0 \times 10^{9}$ M^{-1} and $K_a = 2.0 \times 10^{9}$ M^{-1} respectively). Compound **4** showed results similar to compound **2**. The control compound ImPyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**3**) bound the 5'-T**G**TACA-3' site with $K_a = 2.9 \times 10^{11}$ M^{-1} and the three single base pair mismatches at lower affinities (K_a = 1.1-2.9 × 10^{10} M^{-1}).[5]

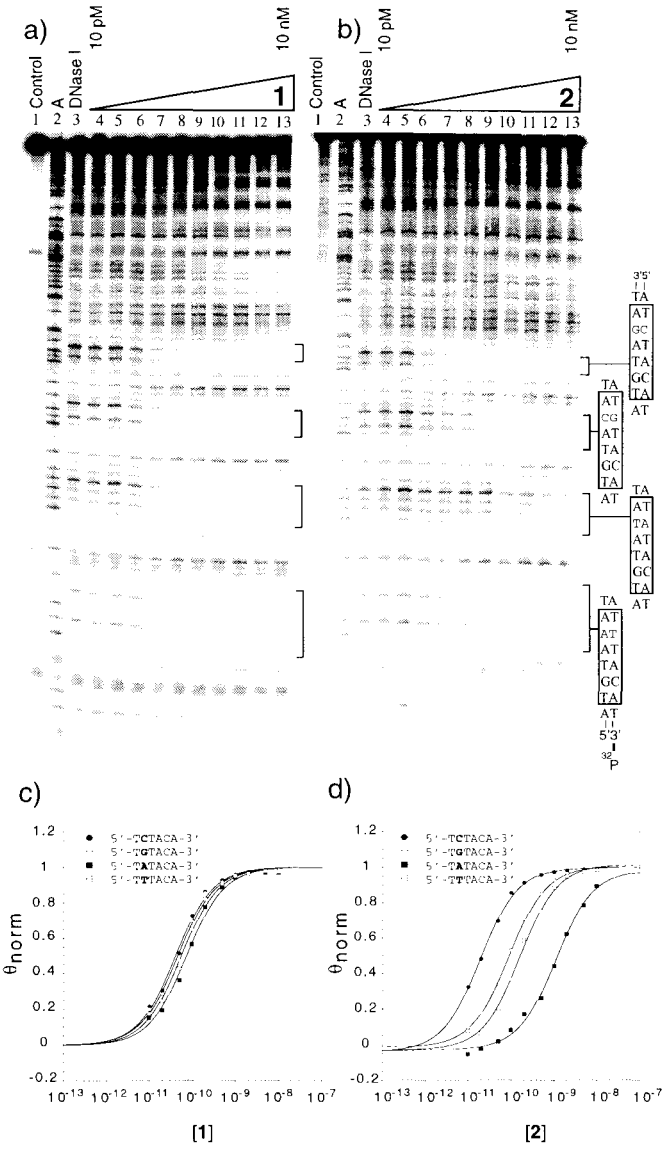


Figure 4.5. (a) Quantitative DNase I footprint titration experiment PyPyPyPy-(R)$H_2N\gamma$-ImPyPyIm-β-Dp (**1**) on the 284 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW15: lane 1, intact DNA; lane 2, A specific reaction; lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM PyPyPyPy-(R)$H_2N\gamma$-ImPyPyIm-β-Dp (**1**); The four sites 5'-TTTACA-3', 5'-TATACA-3', 5'-TGTACA-3' and 5'-TCTACA-3' sites that were analyzed are shown on the right side of the gel. (b) Quantitative DNase I footprint titration PyPyPyPy-(R)$H_2N\gamma$-ImPy-β–Im-β-Dp (**2**) on the 284 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW15: lane 1, intact DNA; lane 2, A specific reaction lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM PyPyPyPy-(R)$H_2N\gamma$-ImPy-β–Im-β-Dp (**2**); The four sites 5'-TTTACA-3', 5'-TATACA-3', 5'-TGTACA-3' and 5'-TCTACA-3' sites that were analyzed are shown on the right side of the gel. c) Binding isotherms derived from the DNase I footprinting gels of polyamide **1** and the four different sites 5'-TTTACA-3', 5'-TATACA-3', 5'-TGTACA-3' and 5'-TCTACA-3'. d) Binding isotherms derived from the DNase I footprinting gels of polyamide **2** and the four different sites 5'-TTTACA-3', 5'-TATACA-3', 5'-TGTACA-3' and 5'-TCTACA-3'.

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

Polyamide	5'-TTTACA-3'	5'-TATACA-3'	5'-TGTACA-3'	5'-TCTACA-3'
1	$1.7\ (\pm0.3) \times 10^{10}$	$1.3\ (\pm0.4) \times 10^{10}$	$1.9\ (\pm0.5) \times 10^{10}$	$1.9\ (\pm0.5) \times 10^{10}$
2	$8.0\ (\pm2.0) \times 10^{9}$	$2.0\ (\pm1.0) \times 10^{9}$	$1.0\ (\pm0.4) \times 10^{10}$	$4.9\ (\pm0.9) \times 10^{10}$
3	$2.0\ (\pm0.5) \times 10^{10}$	$1.1\ (\pm0.2) \times 10^{10}$	$2.9\ (\pm0.5) \times 10^{11}$	$2.9\ (\pm0.5) \times 10^{10}$
4	$1.2\ (\pm0.1) \times 10^{10}$	$6.7\ (\pm4.2) \times 10^{9}$	$1.4\ (\pm0.1) \times 10^{10}$	$4.9\ (\pm1.8) \times 10^{10}$

[a]Values reported are mean values from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. The assays were perfomed at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

Discussion

PyPyPyPy-(R)H_2Nγ-ImPyPyIm-β-Dp (**1**) showed no preference for its designed matched binding site 5'-T**C**TACA-3' over the three single base pair mismatches. The lack of discrimination could be due to the mispositioning of the Im residue located at the C-terminal end of the four ring polyamide subunit. PyPyPyPy-(R)H_2Nγ-ImPy-β–Im-β-Dp (**2**) instead showed a 5-25-fold selectivity for the 5'-T**C**TACA-3' site over the three single base pair mismatched sites. The replacement of a pyrrole ring with a flexible β-alanine resets the imidazole ring to provide the optimal hydrogen bond formation between the imidazole N3 and the exocyclic amine of guanine. This study opens the opportunity to target sequences starting at a C•G base pair and also validates the use of flexible β-alanine residue to reset the register of the C-terminal imidazole amino acid. These results gives us an opportunity to target new biologically important DNA sequences.

Experimental Section

Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-PAM-Resin), N,N'-Dicyclohexylcarbodiimide (DCC), N-Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and Boc-β-alanine were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA) and *N,N*-dimethylformamide (DMF), were purchased from Applied Biosystems. DMF was distilled under reduced pressure prior to synthesis. (R)-2-Fmoc-4-Boc-diaminobutyric acid was from Bachem, dichloromethane (DCM) was reagent grade from EM, thiophenol (PhSH) and dimethylaminopropylamine from Aldrich, and trifluoroacetic acid (TFA) from Halocarbon. ^{1}H NMR were recorded on a Varian Mercury 300 instrument. Chemical

shifts are reported in ppm downfield from the signal for Me_4Si, with reference 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 on a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5μm, 300 × 4.6 mm reversed-phase column in 0.1% (w/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 × 100 mm, 100μm C_{18} column, 0.1% (w/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.

PyPyPyPy-(R)^{Fmoc}γ-ImPyPyIm-β-PAM-resin and PyPyPyPy-(R)^{Fmoc}γ-ImPy-β–Im-β-PAM-resin PyPyPyPy-(R)^{Fmoc}γ-ImPyPyIm-β-PAM-resin (0.333 mmol/g) and PyPyPyPy-(R)^{Fmoc}γ-ImPy-β–Im-β-PAM-resin (0.338 mmol/g) were synthesized in a stepwise fashion from 0.55 mmol/g Boc-β-PAM-resin by manual solid-phase methods.[6] (R)-2-Fmoc-4-Boc-diaminobutyric acid was incorporated as previously described for Boc-γ-aminobutyric acid.[3c,6]

Procedure for cleavage from the resin. After the coupling was completed the resin was filtered off the reaction mixture and washed with DMF (2 × 30 seconds). DMF (1 mL) and piperidine (3 mL) were added and the mixture shaken for 30 minutes at 22°C. The resin was filtered off and washed with DMF (2 × 30 seconds), DCM (3 × 30 seconds), MeOH (1 × 30 seconds), Et_2O (1 × 30 seconds) and dried *in vacuo*. The resin was treated with (dimethylamino)propylamine (1 mL) with periodic agitation at 55°C for 16 hours. The reaction mixture was then filtered to remove the resin, TFA (0.1% (w/v), 7 mL) added and

the resulting solution purified by reversed-phase HPLC. The pure compounds were recovered as white powders upon lyophilization of the appropriate fractions.

PyPyPyPy-(R)^{H_2N}γ-ImPyPyIm-β-Dp (1). PyPyPyPy-(R)^{Fmoc}γ-ImPyPyPy-β-PAM-resin (100 mg) was cleaved and purified by preparative HPLC to give **1** (13 mg, 31% isolated yield). UV λ_{max} (H_2O) 240, 312 (69500); ^{1}H NMR (DMSO-d_6) δ 10.97, 10.26, 10.07 (s, 1 H each), 9.84-9.92 (m, 3 H), 9.78 (s, 1 H), 8.20-8.30 (m, 3 H), 8.12 (t, 1 H, J = 5.8 Hz), 8.03 (t, 2 H, J = 6.3 Hz), 7.91 (t, 1 H, J = 5.8 Hz), 7.48 (s, 1 H), 7.44 (s, 1 H), 7.20 (d, 1 H, J = 1.2 Hz), 7.19 (d, 1 H, J = 1.8 Hz), 7.16 (t, 2 H, J = 1.5 Hz), 7.12 (d, 2 H, J = 1.8 Hz), 7.06 (d, 1 H, J = 1.8 Hz), 7.02 (d, 1 H, J = 1.8 Hz), 6.99 (d, 1 H, J = 1.8 Hz), 6.90 (d, 1 H, J = 1.8 Hz), 6.89 (t, 1 H, J = 2.1 Hz), 6.86 (dd, 1 H, J = 2.1, 4.2 Hz), 5.99 (dd, 1 H, J = 2.7, 4.2 Hz), 3.92, 3.87, 3.81, 3.80, 3.79, 3.78, 3.77, 3. 75 (s, 3 H each), 3.04 (q, 2 H, J = 6.3 Hz), 2.90-3.00 (m, 2 H), 2.67 (d, 6 H, J = 4.8 Hz), 2.30 (t, 2 H, J = 6.0 Hz), 1.90-2.00 (m, 2 H), 1.66 (bp, 2 H, J = 7.8 Hz); MALDI-TOF-MS calcd. for $C_{58}H_{73}N_{22}O_{10}$ (M+H): 1237.6; found: 1237.7.

PyPyPyPy-(R)^{H_2N}γ-ImPy-β–Py-β-Dp (2). PyPyPyPy-(R)^{Fmoc}γ-ImPy-β–Im-β-PAM-resin (100 mg) was cleaved and purified by preparative HPLC to give **2** (8 mg, 18% isolated yield). UV λ_{max} (H_2O) 240, 312 (60830); ^{1}H NMR (DMSO-d_6) δ 11.10, 10.44, 10.13, 10.03, 10.01, 9.92 (s, 1 H each), 8.40, (bs, 2 H), 8.26 (t, 1 H, J = 6.0 Hz), 8.16 (t, 2 H, J = 6.0 Hz), 8.04 (t, 1 H, J = 5.7 Hz), 7.60 (s, 1 H), 7.48 (s, 1 H), 7.30 (t, 2 H, J = 1.5 Hz), 7.29 (d, 1 H, J = 1.8 Hz), 7.25 (d, 1 H, J = 1.5 Hz), 7.15 (d, 1 H, J = 1.5 Hz), 7.13 (d, 1 H, J = 1.5 Hz), 7.04 (d, 1 H, 1.5 Hz), 7.00-7.15 (m, 2 H), 6.99 (dd, 1 H, J = 2.1, 4.2 Hz), 6.13 (dd, 1 H, J = 2.4, 3.9 Hz), 4.04, 3.98, 3.95, 3.93, 3.92, 3.88, 3.87 (s, 3 H each), 3.30-3.40 (m, 2 H), 3.17 (q, 2 H, J = 6.0 Hz), 3.07 (p, 2 H, J = 4.8 Hz), 2.81 (d, 6 H, J = 5.1 Hz), 2.63 (d, 2 H, J = 6.9 Hz), 2.43 (t, 2 H, J = 6.9 Hz), 2.00-2.10 (m, 2 H), 1.80 (p, 2 H, J = 7.2 Hz); MALDI-TOF-MS calcd. for $C_{55}H_{72}N_{21}O_{10}$ (M+H): 1186.6; found: 1186.7.

BzPyPyPy-(R)^{H_2N}γ-ImPyβPy-β-Dp (4). A mixture of DMF (1 mL), DIEA (0.5 mL) and benzoyl chloride (0.05 mL, 0.4 mmol) was added to deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyβIm-β-PAM-resin (100 mg) and shaken for 10 minutes at 22°C. The compound was then cleaved and purified by preparative HPLC to give BzPyPyPy-(R)^{H_2N}γ-ImPyβPy-β-Dp (**4**) (8 mg, 17% isolated yield). UV λ_{max} (H_2O) 240, 312 (60830); ^{1}H NMR (DMSO-d_6) δ 10.97, 10.31, 10.30, 10.00, 9.94, 9.89 (s, 1 H each), 9.29 (bs, 1 H), 8.27 (bs, 2 H), 8.13 (t, 1 H, J = 5.7 Hz), 8.04 (t, 2 H, J = 6.0 Hz), 7.85-7.95 (m, 3 H), 7.40-7.60 (m, 5 H), 7.34 (s, 1 H), 7.27 (d, 1 H, J = 1.8 Hz), 7.18 (d, 1 H, J = 2.1 Hz), 7.16 (d, 1 H, J = 1.8 Hz), 7.12 (d, 1 H, J = 1.8 Hz), 7.06 (d, 1 H, J = 1.8 Hz), 7.02 (d, 1 H, J = 1.8 Hz), 6.91 (d, 1 H, J = 1.8 Hz), 6.88 (d, 1 H, J = 1.8 Hz), 3.90, 3.84, 3.82, 3.80, 3.79, 3.75, 3.74 (s, 3 H each), 3.36 (bp, 2 H, J = 6.0 Hz), 3.22 (q, 2 H, J = 6.3 Hz), 3.03 (q, 2 H,J = 6.3 Hz), 2.93 (p, 2 H, J = 5.7 Hz), 2.67 (d, 6 H, J = 4.8 Hz), 2.49 (t, 2 H, J = 7.2 Hz), 2.29 (t, 2 H, J = 6.6 Hz), 1.90-2.00 (m, 2 H), 1.66 (p, 2 H, J = 7.8 Hz); MALDI-TOF-MS calcd. for $C_{56}H_{71}N_{20}O_{10}$ (M+H): 1183.6; found: 1183.8.

Preparation of 3′- and 5′-End-Labeled Restriction Fragments. The plasmid pCW15[5] was linearized with *Eco*RI and *Pvu*II and then treated with Sequenase (version 2.0 from United States Biochemical), deoxyadenosine 5′-[α-^{32}P]triphosphate and thymidine 5′-[α-^{32}P]triphosphate for 3′ labeling. The 3' labeled fragment was loaded onto a 7% non-denaturing polyacrylamide gel, and the desired 284 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.[7]

Quantitative DNase I Footprinting. DNase I Footprinting reactions were carried out as previously described.[3b] Photostimulable storage phosphorimaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22°C for 12-16 hours. A Molecular

Dynamics Typhoon PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integration of all bands using ImageQuant v. 3.2 software. All DNA manipulations were performed according to standard protocols.[8]

Acknowledgement

We are grateful to the National Institutes of Health (GM-27681) for research support, the Knut and Alice Wallenberg Foundation for a post-doctoral fellowship to U.E. and the National Institutes of Health for a research traineeship award (GM-08501) to C.C.C.W. We thank G. M. Hathaway for MALDI-TOF mass spectrometry.

References

1. Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688-693.
2. (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988 (b) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561. (c) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.
3. White, S., Baird, E. E., Dervan, P. B. *Chem. Biol.* **1997**, ,*4*, 569-578.
4. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.
5. Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, 122, 9354-9360.
6. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
7. (a) Iverson, B. L.; Dervan, P. B. *Nucl. Acids Res.* **1987**, *15*, 7823-7830. (b) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499-560.
8. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.

Chapter 5

Incorporation of β-alanine to Restore Sequence Specificity in the Minor Groove of DNA by Hairpin Polyamides; Discrimination of 5'-CWWG-3' Sequences

Abstract: Eight-ring hairpin polyamides containing three aromatic amino acids, N-methylpyrrole (Py), N-methylimidazole (Im) and 3-hydroxy-1-methylpyrrole (Hp) recognize predetermined six-base pair sites in the minor groove of DNA. ImPyPyIm-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**1**) showed twofold preference for its designed matched binding site 5'-tAGTAGt-3' over the single base pair mismatch 5'-tAGTACt-3' while polyamide ImPy-β-Im-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**2**) instead showed a six-fold selectivity for the 5'-tAGTAGt-3 site over the mismatched site. These results follow earlier findings where replacement of a pyrrole ring with a flexible β-alanine resets the imidazole ring to provide the optimal hydrogen bond formation between the imidazole N3 and the exocyclic amine of guanine. This information was used to design polyamides to target a general 5'-CWWG-3' site. Polyamide PyPy-β-Im-(R)^{H_2N}γ-PyPy-β-Im-β-Dp (**6**), where two Py residues were replaced by β-alanine, differentiated the 5'-tACTAGTg-3' site from the two mismatch sites 5'-tACTACTg-3' and 5'-tAGTAGTg-3'.

This study opens the opportunity to target general 5'-CWWG-3' sequences and validates the use of flexible β-alanine residue to reset the register of adjacent Im residues within the hairpin motif and provides an opportunity to target new biologically important DNA sequences in gene regulation studies.

Introduction

Polyamides containing three aromatic amino acids, N-methylpyrrole (Py), N-methylimidazole (Im) and 3-hydroxy-1-methylpyrrole (Hp) are synthetic ligands that have affinities and specificities for DNA comparable to those of naturally occurring DNA-binding proteins.[1] DNA recognition depends on side by side antiparallel pairing of amino acids in the minor groove. Antiparallel pairing of Im opposite Py (Im/Py) pair binds G•C, while a Py/Im pair recognizes C•G. A Py/Py pair binds both A•T and T•A base pairs in preference to G•C and C•G. An Hp/Py pair prefers T•A over A•T, C•G or G•C. Eight-ring pyrrole imidazole polyamides have been shown to be cell permeable and to inhibit transcription of designated genes in cell culture.[2]

Efforts have been made to increase the DNA-binding affinity and sequence specificity by covalently linking polyamide heterodimers and homodimers. Polyamide dimers linked by a γ-aminobutyric acid unit (γ) creates a hairpin structure which binds with 100-fold enhanced affinity relative to the unlinked subunits. The placement of an Im and Py in the first and eighth position of an eight-ring polyamide creates an Im/Py pair at the N- and C-terminal positions of the hairpin, which targets DNA sequences beginning with a G•C base pair. In the process of designing polyamides capable of targeting sequences beginning with C•G base pairs, we found it was necessary to incorporate a flexible β-alanine monomer to reset the imidazole at the C-terminus of the four-ring subunit.[3]

Numerous eight-ring hairpin polyamides have been synthesized and published by our research group; however, notably absent in the list are compounds containing imidazoles immediately preceding the γ turn due to the lack of specificity of the imidazole ring. To solve this issue it might be necessary to incorporate flexible β-alanine monomer similar to the case for sequences beginning with C•G base pairs. Two head to tail linked hairpin polyamides ImPyPyIm-(R)H_2Nγ-PyPyPyPy-β-Dp (**1**) and ImPy-β–Im-

β-(R)H_2Nγ-PyPyPyPy-β-Dp (**2**) were synthesized and tested towards the general DNA binding sites 5'-CWWC-3' and 5'-GWWG-3'. Compounds **1** and **2** differs from the two previously synthesized polyamides designed for sequences beginning with C•G base pairs PyPyPyPy-(R)H_2Nγ-ImPyPyIm-β-Dp (**3**) and PyPyPyPy-(R)H_2Nγ-ImPy-β–Im-β-Dp (**4**) in the location of the γ turn (Figure 5.1).

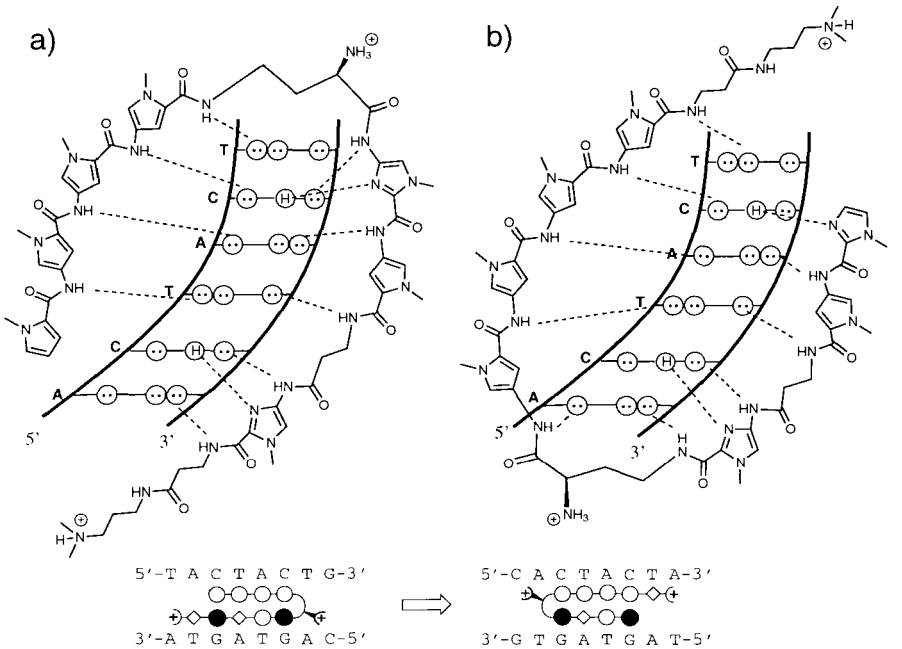


Figure 5.1. (a) Hydrogen bonding model of the polyamide:DNA complex between an seven-ring hairpin polyamide PyPyPyPy(R)H_2Nγ-ImPy-β-Im-β-Dp (**4**) and a 5'-TCTACA-3' site. A circle with two dots represents the lone pairs of N3 of purines and O2 of cytosine. Two touching circles with dots represent the two lone pairs of the O2 of thymine. Circles containing an H represent the exocyclic amino hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. The filled and unfilled circles represent imidazole and pyrrole rings, respectively, curved line represents γ-aminobutyric acid, and diamond represent β-alanine. (b) Hydrogen bonding model of the polyamide:DNA complex between a six-ring hairpin polyamide ImPy-β–Im(R)H_2Nγ-PyPyPyPy-β-Dp (**2**) and a 5'-TCTACA-3' site.

The two polyamides PyPyPyIm-(R)H_2Nγ-PyPyPyIm-β-Dp (**5**) and PyPy-β–Im-β-(R)H_2N γ-PyPy-β-Im-β-Dp (**6**) were then designed to target a general 5'-CWWG-3' site (Figure 5.2). A control compound ImPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**7**) was also synthesized to establish a baseline with previously published work.[4]

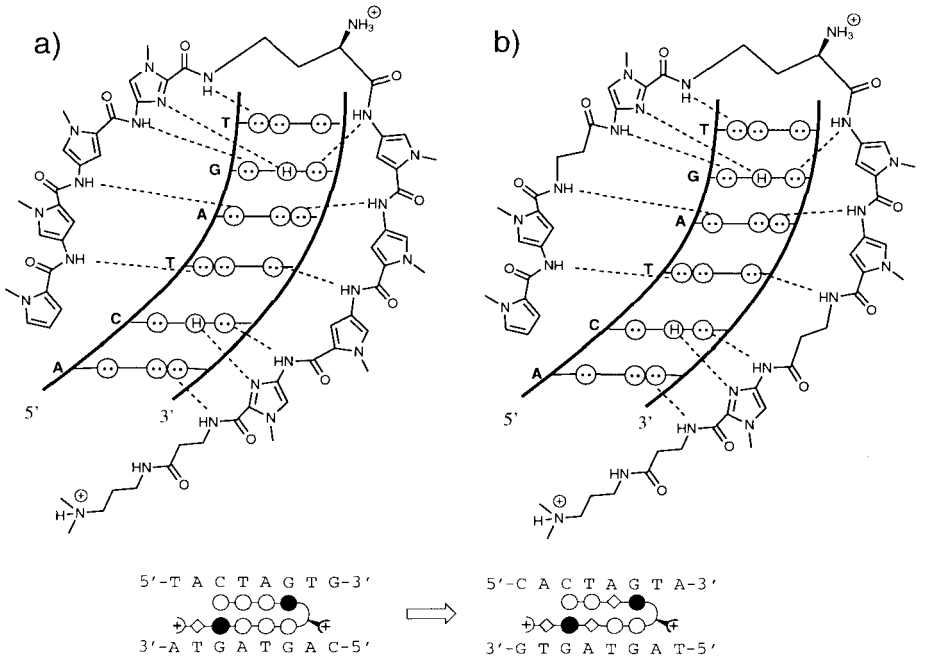


Figure 5.2. (a) Hydrogen bonding model of the polyamide:DNA complex between an eight-ring hairpin polyamide PyPyPyIm-(R)H_2Nγ-PyPyPyIm-β-Dp (**5**) and a 5'-TCTA**G**A-3' site. (b) Hydrogen bonding model of the polyamide:DNA complex between a seven-ring hairpin polyamide PyPy-β–Im-(R)H_2Nγ-PyPy-β–Im-β-Dp (**4**) and a 5'- TCTA**G**A -3' site.

The plasmid pCW20 was designed to contain the four six-base pair recognition sites 5'-tA**C**TA**G**Tg-3', 5'-tA**C**TA**C**Tg-3', 5'-tA**G**TA**G**Tg-3', and 5'-tA**G**TA**C**Tg-3' which differ by the placement of the cytosine and guanine bases. We chose different flanking sequences bordering the core four base pairs in order to bias the binding of the polyamide to only one direction. In one direction the beta-tail would be over the guanine base pair which has been shown to be deleterious for binding.[5]

Results and Discussion

Synthesis: Polyamides **3**, **4**, and **7** were prepared as previously described.[3,4] ImPyPyIm-(R)^{Fmoc}γ-PyPyPyPy-β-PAM-resin, ImPy-β–Im-β-(R)^{Fmoc}γ-PyPyPyPy-PAM-resin, PyPyPyIm-(R)^{Fmoc}γ-PyPyPyIm-β-PAM-resin and PyPy-β–Im-(R)^{Fmoc}γ-PyPy-β–Im-β-

PAM-resin were synthesized in a stepwise manner from Boc-β-alanine-PAM resin (0.55 mmol/g) using solid-phase methodology[6] in 14 steps. A sample of the resin was then cleaved by aminolysis with ((dimethyl)amino)propylamine (55°C, 16 hours) and purified by reversed-phase HPLC to provide ImPyPyIm-(R)H_2Nγ-PyPyPyPy-β-Dp (**1**), ImPy-β–Im-β-(R)H_2Nγ-PyPyPyPy-β-Dp (**2**), PyPyPyIm-(R)H_2Nγ-PyPyPyIm-β-Dp (**5**), and PyPy-β–Im-(R)H_2Nγ-PyPy-β–Im-β-Dp (**6**) (Figure 5.3). For the synthesis of compounds **1-7-E**, a sample of the parent polyamide was treated with an excess of EDTA-dianhydride (DMSO/NMP, DIEA 55°C, 30 minutes), and the remaining anhydride was hydrolyzed (0.1M NaOH, 55°C, 10 minutes). The polyamide-EDTA conjugates **1-7-E** were then isolated by reversed-phase HPLC.

Identification of binding orientation by affinity cleaving: Affinity cleavage experiments with compounds **1-7-E,** which have an EDTA•Fe(II) moiety attached to the γ-turn, were used to confirm binding orientation. Affinity cleavage titration were performed on the 3'- or 5'-^{32}P-end labeled *Eco*RI/*Pvu*II restriction fragment from the plasmid pCW20 which contains the four binding sites 5'-tACTAGTg-3', 5'-tACTACTg-3', 5'-tAGTAGTg-3', and 5'-tAGTACTg-3'. The affinity cleavage assays revealed cleavage patterns that are 3'-shifted, as is expected for hairpin formation and occupation in the minor groove of DNA. For the two sites 5'-tACTACTg-3' and 5'-tAGTAGTg-3' a single cleavage pattern was observed for compounds **1-4-E**, consistent with hairpin polyamides binding in a forward orientation (5'-3', N-C) (Figure 5.4). We observed that polyamides **1**- **4-E** bound the two sites 5'-tAGTACTg-3' and 5'-tACTAGTg-3' in only one orientation. The cleavage patterns were located on the right side of the binding site suggesting a binding orientation which would place the Dp charge over an A•T base pair instead of G•C. A single cleavage pattern 3' for all four binding sites was observed for affinity compounds **5-E** and **6-E** (Figure 5.5). For compound **7-E** single cleavage pattern was observed for the DNA sites 5'-tACTACTg-3', 5'-tAGTACTg-3' and 5'-tACTAGTg-3'

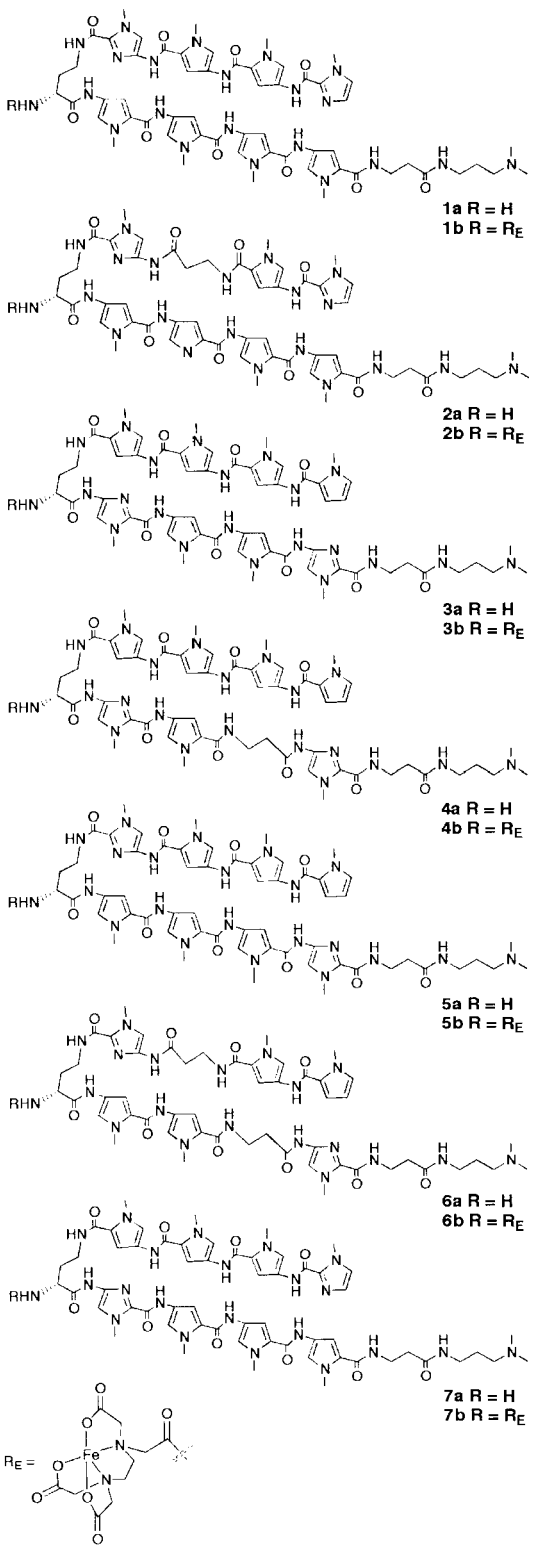


Figure 5.3. Structures of polyamides and their respective affinity cleavage analogs ImPyPyIm-(R)H_2Nγ-PyPyPyPy-β-Dp (**1**), ImPy-β–Im-β-(R)H_2Nγ-PyPyPyPy-β-Dp (**2**), PyPyPyPy-(R)H_2Nγ-ImPyPyIm-β-Dp (**3**), PyPyPyPy(R)H_2Nγ-ImPy-β-Im-β-Dp (**4**), PyPyPyIm-(R)H_2Nγ-PyPyPyIm-β-Dp (**5**), PyPy-β–Im-(R)H_2Nγ-PyPy-β–Im-β-Dp (**6**), and ImPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**7**).

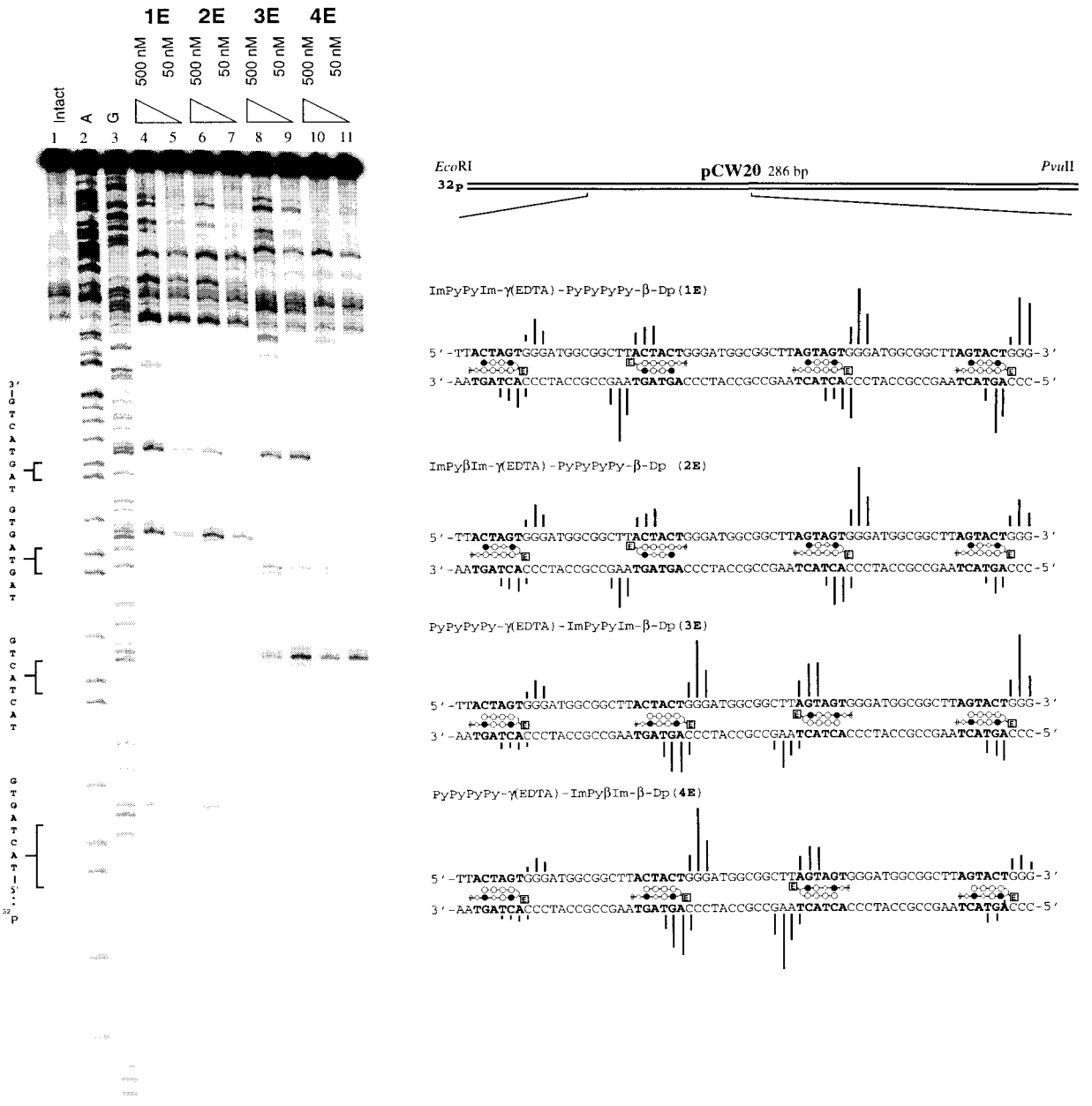


Figure 5.4. Affinity cleavage experiments with EDTA-polyamides **1E-4E** on the 5' ^{32}P-labeled 284 base pair *Eco*RI/*Pvu*II restriction fragment derived from plasmid pCW20. The designed 5'-ACTAGT-3', 5'-ACTACT-3', 5'-AGTAGT-3', and 5'-AGTACT-3' are shown on the left side of the autoradiogram. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4-11, digestion products obtained in the presence of **1-E**, **2-E**, **3-E**, or **4-E**; 500 nM or 50 nM polyamide. Results of affinity cleavage experiments with EDTA-polyamides **1E-4E**. Arrow height is proportional to the amount of cleavage at the indicated base.

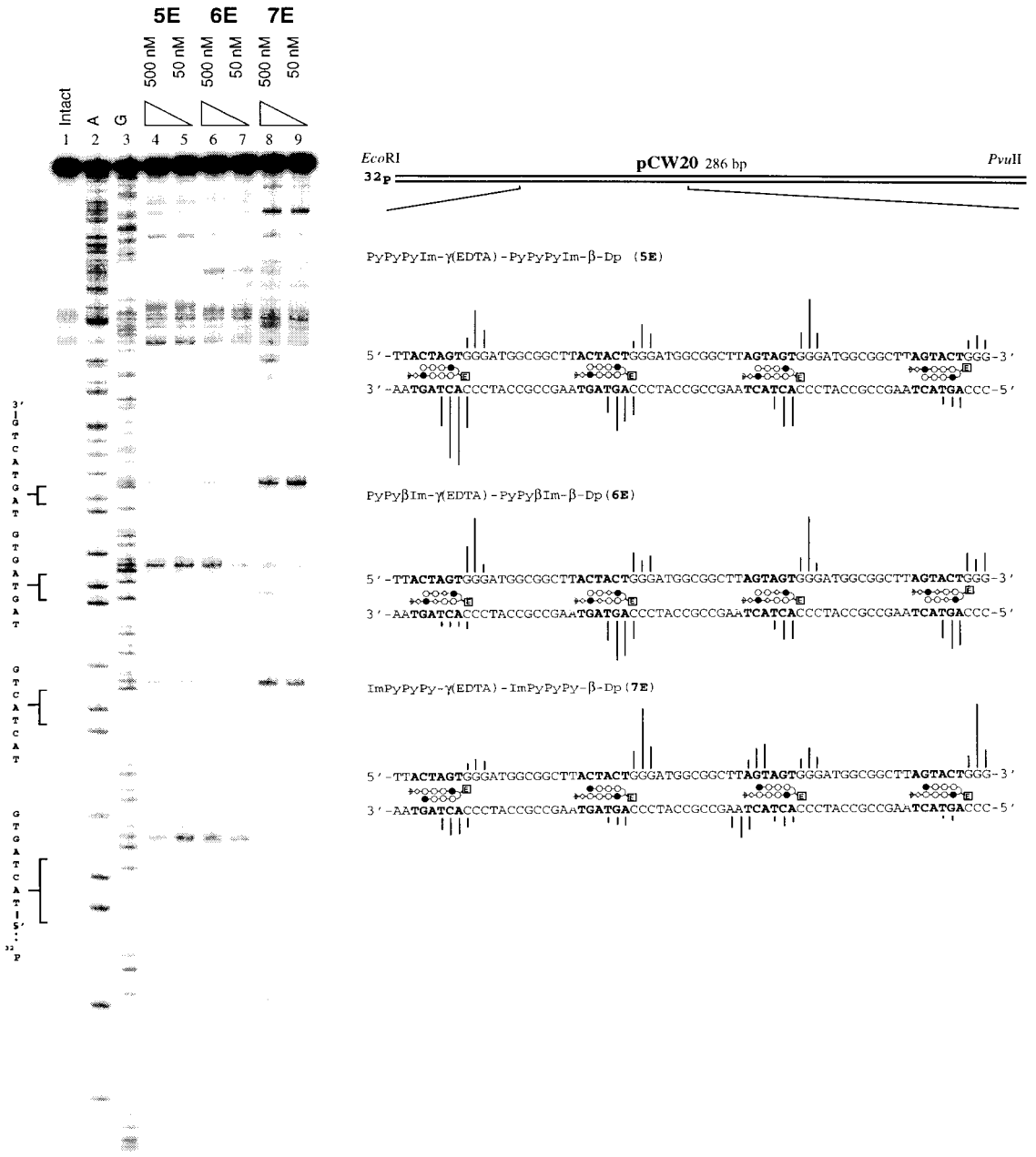


Figure 5.5. Affinity cleavage experiments with EDTA-polyamides **5E-7E** on the 5' ^{32}P-labeled 284 base pair *Eco*RI/*Pvu*II restriction fragment derived from plasmid pCW20. The designed 5'-ACTAGT-3', 5'-ACTACT-3', 5'-AGTAGT-3', and 5'-AGTACT-3' are shown on the left side of the autoradiogram. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4-11, digestion products obtained in the presence of **5-E**, **6-E**, or **7-E**; 500 nM or 50 nM polyamide. Results of affinity cleavage experiments with EDTA-polyamides **5E-7E**. Arrow height is proportional to the amount of cleavage at the indicated base.

and two cleavage patterns was observed for 5'-tAGTAGTg-3'. For compounds **5-E** and **6-E** to bind to the 5'-tACTAGTg-3' site while following the pairing rules would require the hairpin polyamides to be binding in a forward orientation (5'-3', N-C) while the two polyamides need to bind in a reverse orientation (3'-5', N-C) to the 5'-tAGTACTg-3' site. For polyamide **7-E**, 5'-tACTAGTg-3' would instead be a reverse site and 5'-tAGTACTg-3' would be a forward site.

Equilibrium association constants: Quantitative DNase I footprint titrations (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$, pH 7.0, 22°C) (Figure 5.6) were performed on the 5'-^{32}P-end-labeled 286 bp PCR fragment from plasmid pCW20 to determine the equilibrium association constants for polyamide **1-7** for the four target sites (5'-tACTAGTg-3', 5'-tACTACTg-3', 5'-tAGTAGTg-3', 5'-tAGTACTg-3'). Polyamide **1** and **2** bound their designed match site 5'-tAGTAGTg-3' with comparable affinities K_a = 7.9 x 10^9 M^{-1} and K_a = 8.6 x 10^9 M^{-1}. For the 5'-tAGTACTg-3' site, polyamide **1** bound with an affinity of K_a = 3.8 x 10^9 M^{-1} and polyamide **2** bound with an affinity of K_a =1.3 x 10^9 M^{-1}. We observed greater specificity between the 5'-tAGTAGTg-3' and 5'-tAGTACTg-3' sites for compound **2**, probably due to the flexible β alanine. Polyamide **1** and **2** bound 5'-tACTAGTg-3' with an affinity of K_a =1.6 x 10^9 M^{-1} and K_a =1.7 x 10^9 M^{-1} respectively. In both cases a five-fold drop in affinity compared to the match site was observed due to the selectivity of the Im/Py pair when Im is at the N-terminus of the ring system. The mismatch site 5'-tACTACTg-3' was bound by both **1** and **2** with three to five-fold lower affinity compared to the match site due to the placement of the Dp tail over the G•C base pair. The same trend was observed for polyamide **3** and **4**. The match sites are bound by **3** and **4** with comparable affinities (K_a =1.7 x 10^{10} M^{-1} and K_a =2.0 x 10^{10} M^{-1}). However, polyamide **4** with the flexible β alanine bound 5'-tAGTACTg-3' with twelve-fold lower affinity while **3** showed no preference between the two sites. Polyamide **5** and **6** were designed to bind 5'-tACTAGTg-3'. Polyamide **5** bound 5'-

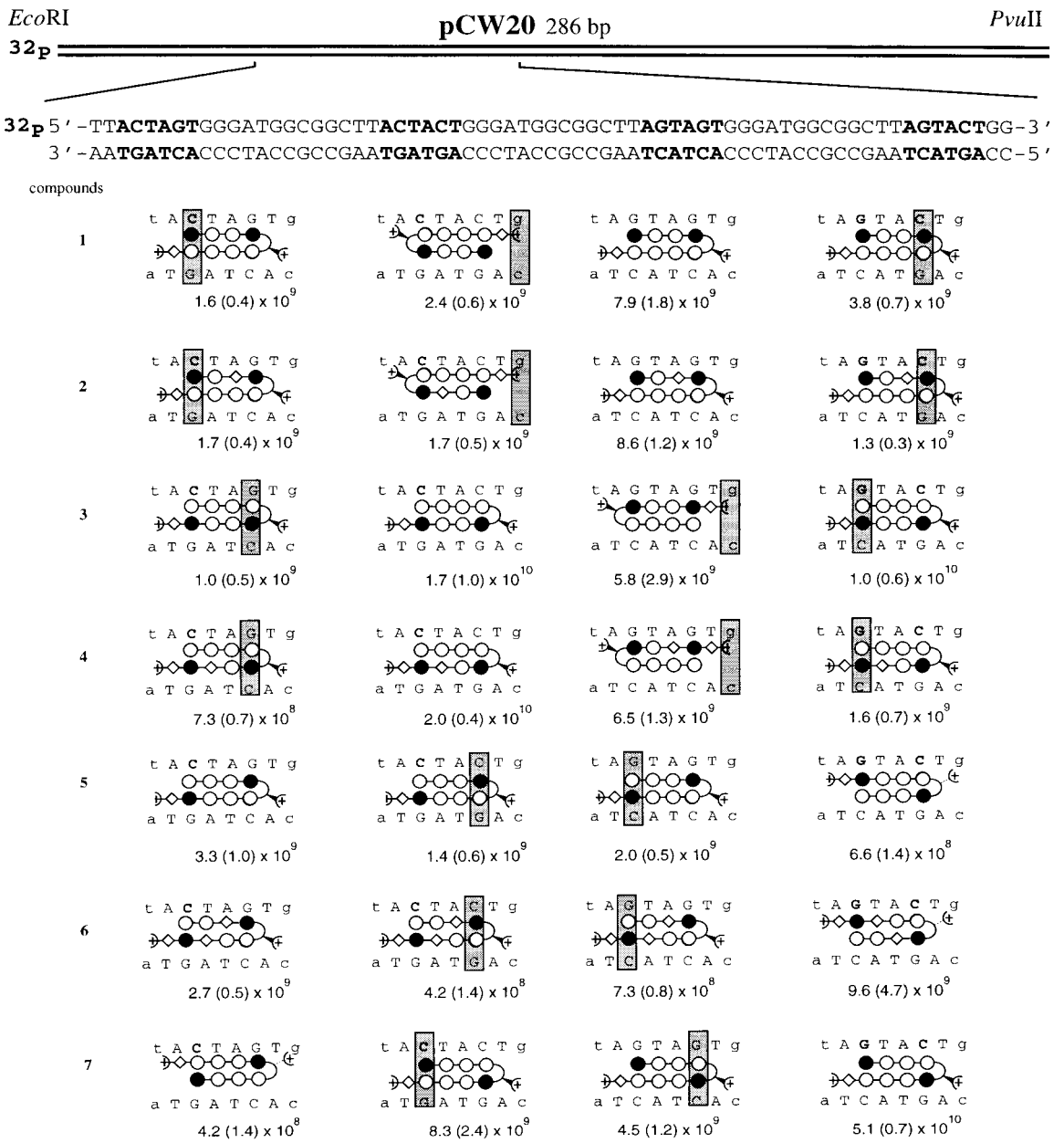


Figure 5.6. Partial sequence of the synthesized insert from pCW20 and the illustration of the EcoRI/PvuII restriction fragment. The sequences in bold were the only sites analyzed by quantitative DNase I footprint titrations. Equilibrium association constants [M^{-1}] and schematic binding models for ImPyPyIm-(R)H_2Nγ-PyPyPyPy-β-Dp (**1**), ImPy-β–Im-β-(R)H_2Nγ-PyPyPyPy-β-Dp (**2**), PyPyPyPy-(R)H_2Nγ-ImPyPyIm-β-Dp (**3**), PyPyPyPy(R)H_2Nγ-ImPy-β-Im-β-Dp (**4**), PyPyPyIm-(R)H_2Nγ-PyPyPyIm-β-Dp (**5**), PyPy-β–Im-(R)H_2Nγ-PyPy-β–Im-β-Dp (**6**), and ImPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**7**) with their putative match and mismatch sites. Values reported are the mean values obtained from three DNase I titration experiments. Mismatch base pairs are marked with a shaded rectangle.

tACTAGTg-3', 5'-tACTACTg-3', and 5'-tAGTAGTg-3' with little specificity K_a =3.3 x $10^9 M^{-1}$, K_a = 1.4 x $10^9 M^{-1}$, and K_a = 2.0 x $10^9 M^{-1}$ while Polyamide **6,** with two β alanine residues, restores the specificity. The two mismatch sites 5'-tACTACTg-3' and 5'-tAGTAGTg-3' are bound with six-fold and four-fold lower affinity compared to the match site 5'-tACTAGTg-3'.

Recognition of 5'-GWWG-3' sequence: Polyamide **1** bound its match site 5'-tAGTAGTg-3' with only twofold higher affinity than its single base pair mismatch site 5'-tAGTACTg-3' suggesting that the Im residue located at the C-terminal end of the four-ring polyamide subunit is not recognizing the exocyclic amine of guanine properly. Polyamide **2** with the flexible β-alanine subunit shows a six-fold specificity between the 5'-tAGTAGTg-3' and 5'-tAGTACTg-3' sites (Figure 5.7). These results suggest the need of the flexible β-alanine residue to restore specificity. Similarly to the results from polyamides **1** and **2**, we observed that polyamide **3** does not recognize 5'-tACTACTg-3' specifically over 5'-tAGTACTg-3'. Incorporation of the flexible β-alanine, as in polyamide **4**, differentiates the two sites (Figure 5.8). Comparison between polyamides **1-4** shows that a flexible β-alanine is necessary to reset the fourth imidazole in the four-ring subunit and the result is general, regardless of the placement of the turn.

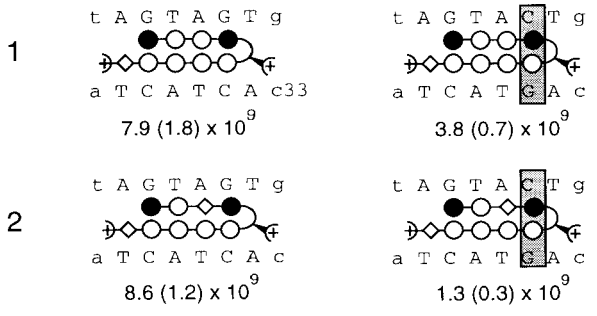


Figure 5.7. Equilibrium association constants [M^{-1}] and schematic binding models for ImPyPyIm-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**1**) and ImPy-β–Im-β-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**2**) with their putative match site 5'-AGTAGT-3' and mismatch site 5'-AGTACT-3'.

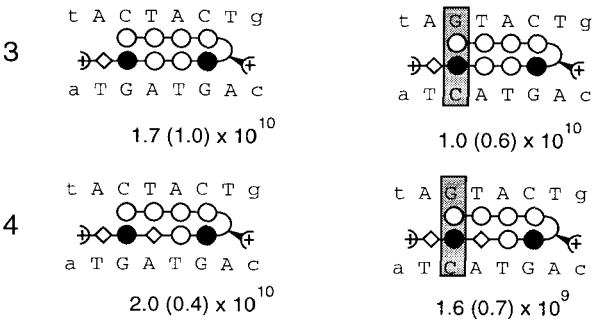


Figure 5.8. Equilibrium association constants [M^{-1}] and schematic binding models for PyPyPyPy-(R)$H_2N\gamma$-ImPyPyIm-β-Dp (**3**) and PyPyPyPy(R)$H_2N\gamma$-ImPy-β-Im-β-Dp (**4**) with their putative match site 5'-ACTACT-3' and mismatch site 5'-AGTACT-3'.

Recognition of 5'-CWWG-3' sequence: The results from polyamide **1-4** demonstrates the necessity of β-alanine to target guanine in the fourth position. Polyamide **5**, without any β-alanine residues, does not differentiate between the 5'-tACTAGTg-3', 5'-tACTACTg-3', and 5'-tAGTAGTg-3' sites. However, placement of two β-alanine residues enables the differentiation of the 5'-tACTAGTg-3' sites from the two mismatch sites (Figure 5.9).

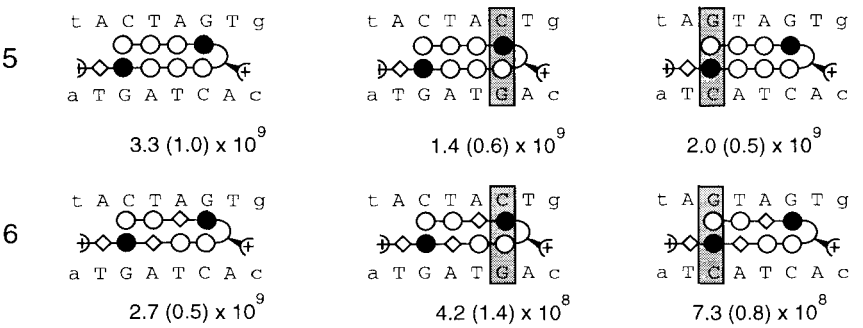


Figure 5.9. Equilibrium association constants [M^{-1}] and schematic binding models for PyPyPyIm-(R)$H_2N\gamma$-PyPyPyIm-β-Dp (**5**) and PyPy-β–Im-(R)$H_2N\gamma$-PyPy-β–Im-β-Dp (**6**) with their putative match site 5'-ACTAGT-3' and mismatch sites 5'-ACTACT-3' and 5'-AGTAGT-3'.

Binding orientation: Quantitative DNase I footprinting experiments show that for the site 5'-tACTAGT-3' compounds **5** and **6** binding in the forward orientation are tighter binders than compound **7** which binds in the reverse orientation. Similarly for the site 5'-tAGTACT-3' compound **7** which binds in a forward orientation to the site binds

tighter than **5** and **6** which binds to the same site in a reverse orientation. We also observe β-alanine containing compounds to bind tighter than their all ring counterparts when binding to DNA sites in a reverse orientation (Figure 5.10).

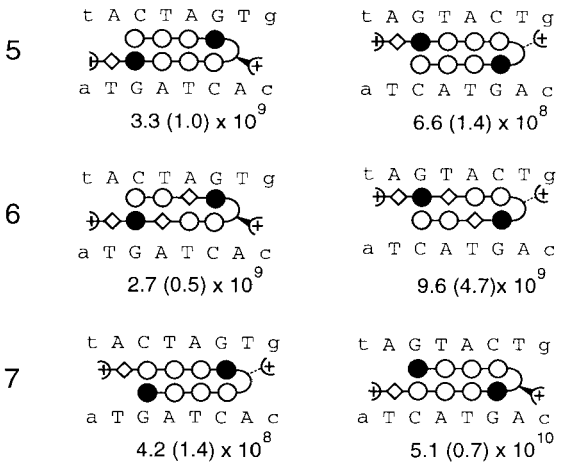


Figure 5.10. Equilibrium association constants [M^{-1}] and schematic binding models for PyPyPyIm-(R)H_2Nγ-PyPyPyIm-β-Dp (**5**), PyPy-β–Im-(R)H_2Nγ-PyPy-β–Im-β-Dp (**6**), and ImPyPyPy-(R)H_2Nγ-ImPyPyPy-β-Dp (**7**) with their putative forward site 5'-ACTAGT-3' and reverse sites 5'-AGTACT-3'.

Conclusion

The classic eight-ring hairpin ImPyPyIm-(R)H_2Nγ-PyPyPyPy-β-Dp (**1**) showed low preference for its designed matched binding site 5'-tAGTAGt-3' over the single base pair mismatch 5'-tAGTACt-3' while polyamide ImPy-β-Im-(R)H_2Nγ-PyPyPyPy-β-Dp (**2**) instead showed a six-fold selectivity for the 5'-tAGTAGt-3 site over the mismatched site. The use of a flexible β-alanine instead of a Py ring resets the imidazole ring to provide the optimal hydrogen bond formation between the imidazole N3 and the exocyclic amine of guanine. These results provide a solution to design hairpin polyamides with imidazole rings preceding a γ turn unit including hairpin polyamides capable in targeting the sequence 5'-CWWG-3'. PyPyPyIm-(R)H_2Nγ-PyPyPyIm-β-Dp (**5**) did not differentiate between the 5'-tACTAGTg-3', 5'-tACTACTg-3', and 5'-tAGTAGTg-3' sites. Polyamide PyPy-β-Im-(R)H_2Nγ-PyPy-β-Im-β-Dp (**6**), where two Py residues were replaced by β-alanine, enabled differentiation of the 5'-tACTAGTg-3' sites from the two mismatch sites.

This study opens the opportunity to target general 5'-CWWG-3' sequences and validates the use of flexible β-alanine residue to reset the register of adjacent Im residues within the hairpin motif and gives us an opportunity to target new biologically important DNA sequences in gene regulation studies.

Experimental Section

General: Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-PAM-Resin), N,N'-Dicyclohexylcarbodiimide (DCC), N-Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and Boc-β-alanine were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA) and *N,N*-dimethylformamide (DMF) were purchased from Applied Biosystems. DMF was distilled under reduced pressure prior to synthesis. (R)-2-Fmoc-4-Boc-diaminobutyric acid was from Bachem, dichloromethane (DCM) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine from Aldrich, and trifluoroacetic acid (TFA) from Halocarbon. ^{1}H NMR were recorded on a Varian Mercury 300 instrument. Chemical shifts are reported in ppm downfield from the signal for Me_4Si, with reference 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 on a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5 μm, 300 × 4.6 mm reversed-phase column in 0.1% (w/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 × 100 mm, 100 μm C_{18} column, 0.1% (w/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. Reagent-grade chemicals were

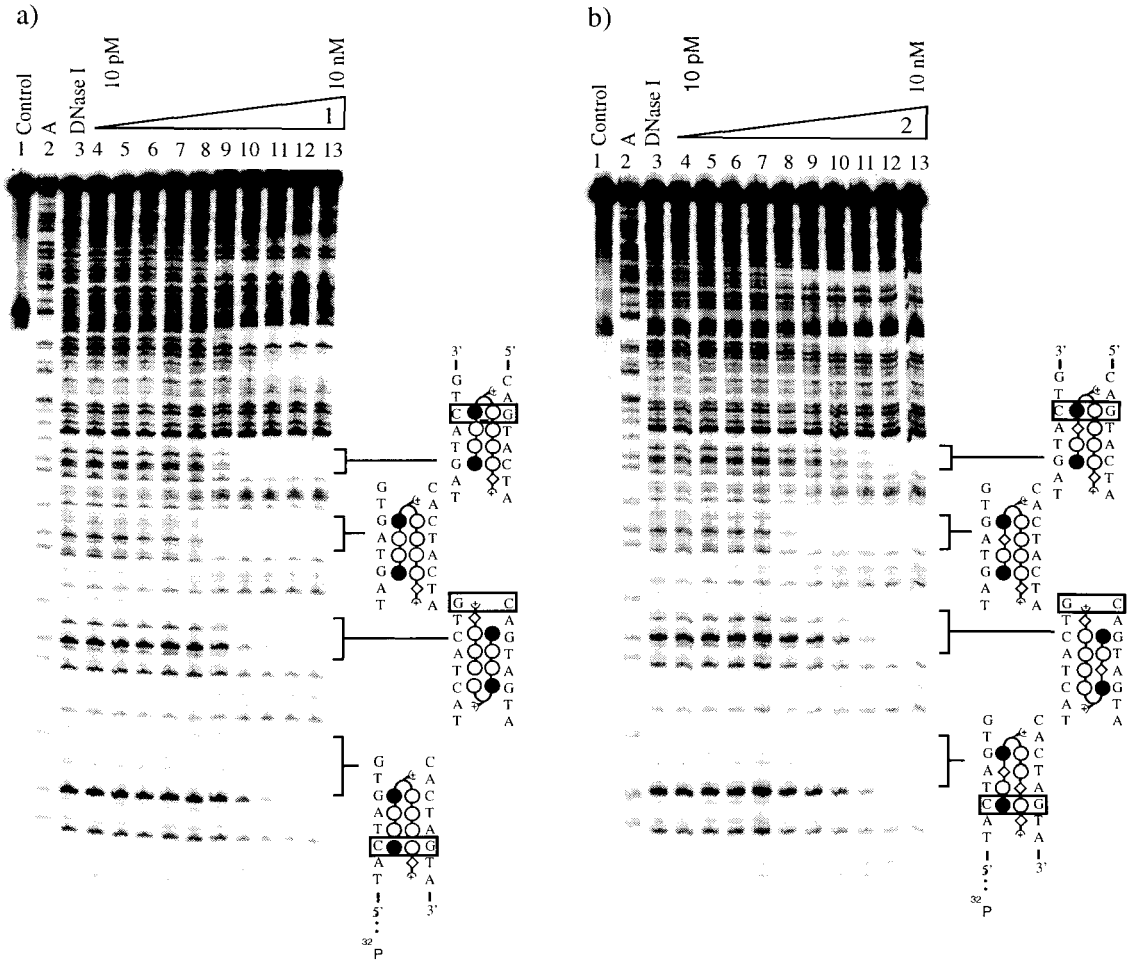


Figure 5.S1. Quantitative DNase I footprint titration experiment for ImPyPyIm-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**1**) on the 286 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW20: lane 1, intact DNA; lane 2, A specific reaction; lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM ImPyPyIm-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**1**). The four sites 5'-T**T**TACA-3', 5'-T**A**TACA-3', 5'-T**G**TACA-3' and 5'-T**C**TACA-3' sites that were analyzed are shown on the right side of the gel. (b) Quantitative DNase I footprint titration for ImPy-β–Im-β-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**2**) on the 286 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW20: lane 1, intact DNA; lane 2, A specific reaction lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM ImPy-β–Im-β-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**2**).

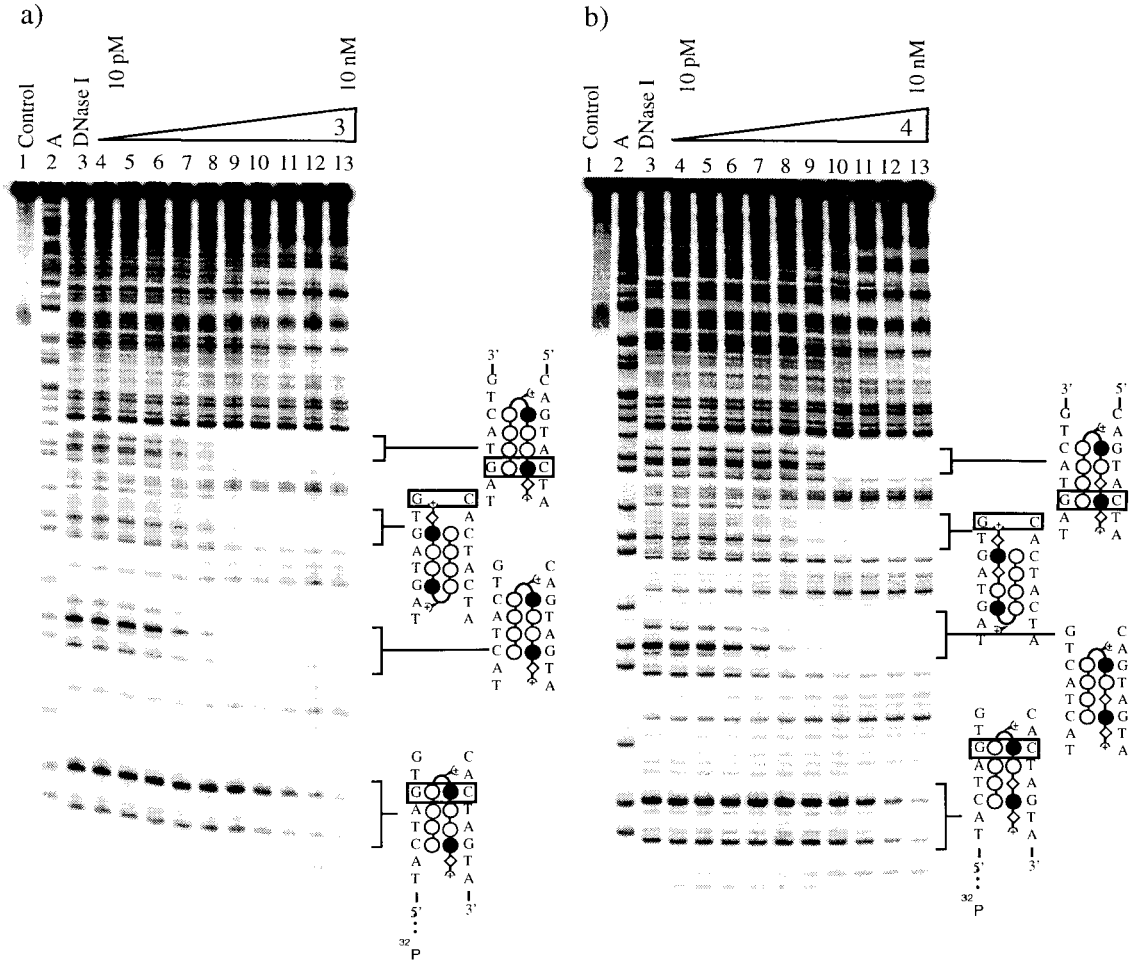


Figure S2. (a) Quantitative DNase I footprint titration experiment for PyPyPyPy-(R)^{H_2N}γ-ImPyPyPyIm-β-Dp (**3**) on the 286 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW20: lane 1, intact DNA; lane 2, A specific reaction; lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM PyPyPyPy-(R)^{H_2N}γ-ImPyPyPyIm-β-Dp (**3**). (b) Quantitative DNase I footprint titration for PyPyPyPy(R)^{H_2N}γ-ImPy-β-Im-β-Dp (**4**) on the 286 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW20: lane 1, intact DNA; lane 2, A specific reaction lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM PyPyPyPy(R)^{H_2N}γ-ImPy-β-Im-β-Dp (**4**).

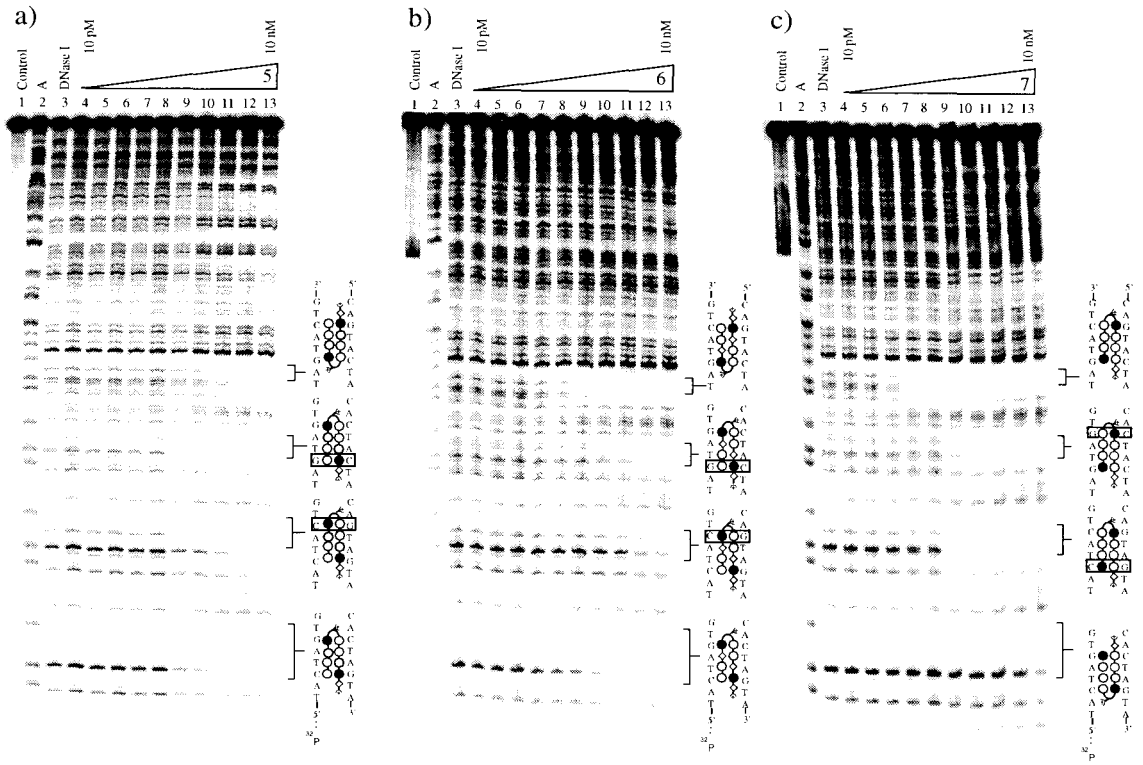


Figure S3. (a) Quantitative DNase I footprint titration experiment for PyPyPyIm-(R)^{H_2N}γ-PyPyPyIm-β-Dp (**5**) on the 286 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW20: lane 1, intact DNA; lane 2, A specific reaction; lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM PyPyPyIm-(R)^{H_2N}γ-PyPyPyIm-β-Dp (**5**). (b) Quantitative DNase I footprint titration for PyPy-β–Im-(R)^{H_2N}γ-PyPy-β–Im-β-Dp (**6**) on the 286 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW20: lane 1, intact DNA; lane 2, A specific reaction lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM PyPy-β–Im-(R)^{H_2N}γ-PyPy-β–Im-β-Dp (**6**). (c) Quantitative DNase I footprint titration for ImPyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**7**) on the 286 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW20: lane 1, intact DNA; lane 2, A specific reaction lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM ImPyPyPy-(R)^{H_2N}γ-ImPyPyPy-β-Dp (**7**).

used unless otherwise stated.

Polyamide synthesis: Reagents and protocols for polyamide synthesis were as previously described.[5] Polyamides were purified by reversed-phase HPLC on a Beckman HPLC with a Waters DeltaPak 25x100 mm, 100 m C18 column equipped with a guard, 0.1% (w/v) TFA, 0.25% acetonitrile min^{-1}. Extinction coefficients were calculated based on ε=8333 per ring at 304 nm.

Procedure for cleavage from the resin: After the coupling was completed, the resin was filtered off the reaction mixture and washed with DMF (2 × 30 seconds). DMF (1 mL) and piperidine (3 mL) were added and the mixture shaken for 30 minutes at 22°C. The resin was filtered off and washed with DMF (2 × 30 seconds), DCM (3 × 30 seconds), MeOH (1 × 30 seconds), Et_2O (1 × 30 seconds) and dried *in vacuo*. The resin was treated with (dimethylamino)propylamine (1 mL) with periodic agitation at 55°C for 16 hours. The reaction mixture was then filtered to remove the resin, TFA (0.1% (w/v), 7 mL) was added and the resulting solution purified by reversed-phase HPLC. The pure compounds were recovered as white powders upon lyophilization of the appropriate fractions.

ImPyPyIm-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**1**): ImPyPyIm-(R)^{Fmoc}γ-PyPyPyPy-β-PAM-resin (100 mg) was cleaved and purified by preparative HPLC to give **1** (5 mg, 3% isolated yield). UV λ_{max} (H_2O) 240, 312 (69500); MALDI-TOF-MS calcd for $C_{58}H_{73}N_{22}O_{10}$ (M+H): 1237.58; found: 1237.7.

ImPy-β-Im-(R)^{H_2N}γ-PyPyPyPy-β-Dp (**2**): ImPy-β-Im-(R)^{Fmoc}γ-PyPyPyPy-β-PAM-resin (100 mg) was cleaved and purified by preparative HPLC to give **2** (15 mg, 8% isolated yield). UV λ_{max} (H_2O) 240, 312 (60830); MALDI-TOF-MS calcd for $C_{55}H_{72}N_{21}O_{10}$ (M+H): 1186.57; found: 1186.7.

PyPyPyIm-(R)^{H_2N}γ-PyPyPyIm-β-Dp (**5**): PyPyPyIm-(R)^{Fmoc}γ-PyPyPyIm-β-PAM-resin (100 mg) was cleaved and purified by preparative HPLC to give **5**

13% isolated yield). UV λ_{max} (H_2O) 240, 312 (69500); MALDI-TOF-MS calcd for $C_{58}H_{73}N_{22}O_{10}$ (M+H): 1237.58; found: 1237.5.

PyPy-β–Im-(R)^{H_2N}γ-PyPy-β-Im-β-Dp (**6**): PyPy-β-Im-(R)^{Fmoc}γ-PyPy-β–Im-β-PAM-resin (100 mg) was cleaved and purified by preparative HPLC to give **6** (14 mg, 11% isolated yield). UV λ_{max} (H_2O) 240, 312 (52160); MALDI-TOF-MS calcd for $C_{55}H_{72}N_{21}O_{10}$ (M+H): 1135.56; found: 1135.7.

Synthesis of EDTA conjugates

Excess EDTA-dianhydride (50 mg) was dissolved in DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 55°C for 5 minutes. The dianhydride solution was added to the hairpin polyamide (10 mg, 5 μmol) dissolved in DMSO (750 μL). The mixture was heated (55°C, 25 minutes) and the remaining EDTA-anhydride hydrolyzed (0.1M NaOH, 3 mL, 55°C, 10 minutes). 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 EDTA conjugate as white powders upon lyophilization of the appropriate fractions

ImPyPyIm-(R)^{EDTA}γ-PyPyPyPy-β-Dp (**1-E**): (0.1 mg, 12% isolated yield). UV λ_{max} (H_2O) 240, 312 (69500); MALDI-TOF-MS calcd for $C_{58}H_{73}N_{22}O_{10}$ (M+H): 1511.66; found: 1511.7.

ImPy-β-Im-(R)^{EDTA}γ-PyPyPyPy-β-Dp (**2-E**): (0.2 mg, 16% isolated yield). UV λ_{max} (H_2O) 240, 312 (60830); MALDI-TOF-MS calcd for $C_{55}H_{72}N_{21}O_{10}$ (M+H): 1460.65; found: 1460.7.

PyPyPyIm-(R)^{EDTA}γ-PyPyPyIm-β-Dp (**5-E**): (0.1 mg, 15% isolated yield). UV λ_{max} (H_2O) 240, 312 (69500); MALDI-TOF-MS calcd for $C_{58}H_{73}N_{22}O_{10}$ (M+H): 1511.66; found: 1511.7.

PyPy-β–Im-(R)^{EDTA}γ-PyPy-β-Im-β-Dp (**6-E**): (1.4 mg, 32% isolated yield). UV λ_{max} (H_2O) 240, 312 (52160); MALDI-TOF-MS calcd for $C_{55}H_{72}N_{21}O_{10}$ (M+H): 1409.64; found: 1409.7.

Construction of Plasmid DNA: The plasmid pcw20 was constructed using previously described methods.[6] Fluorescent sequencing, performed at the DNA Sequencing Facility at the California Institute of Technology, was used to verify the presence of the desired insert.

PCR Labeling to generate 5'-End-Labeled Restriction Fragments. Two 21 mer primers were synthesized for PCR amplification: primer Eco 5'-AATTCGAGCTCGGTACCCGGG-3' and primer Pvu 5'-CTGGCACGACAGGTTTCCCGA-3'. The labeled primer was treated with T4 polynucleotide kinase and deoxyadenosine 5'-[γ-^{32}P] triphosphate as previously described[6] PCR reactions containing 60 pmol each primer, 10 μl PCR buffer (Boehringer-Mannheim), 3.7 μl template (0.003 μg/mL), 2 μl dNTP mix (each at 10 mM), 1 μl 100X BSA (New England Biolabs) and 83 μl water were heated at 70 C for 5 minutes. Four units of Taq Polymerase were added (Boehringer-Mannheim). Thirty amplification cycles were performed, each cycle consisting of the following segments: 94°C for 1 min, 54°C for 1 minute, and 72°C for 1.5 minutes. Following the last cycle, 10 minutes of extension at 72°C completed the reaction. The 5' labeled fragment was loaded onto a 7% non-denaturing polyacrylamide gel, and the desired 286 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.[7]

Quantitative DNase I Footprinting: DNase I Footprinting reactions were carried out as previously described.[6] Photostimulable storage phosphorimaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22°C for 12-16 hours. A Molecular

Dynamics Typhoon PhosphorImager was used to obtain data from the storage screens. The data were analyzed by performing volume integration of all bands using ImageQuant v. 3.2 software. All DNA manipulations were performed according to standard protocols.[8]

Acknowledgement

We are grateful to the National Institutes of Health (GM-27681) for research support, the Knut and Alice Wallenberg Foundation for a post-doctoral fellowship to U.E. and the National Institutes of Health for a research traineeship award (GM 08501) to C.C.C.W. We thank G. M. Hathaway for MALDI-TOF mass spectrometry.

References

[1] P. B. Dervan, R. W. Burli, *Curr. Opin. Chem. Biol* **1999**, *3*, 688-693.

[2] a) J. M. Gottesfeld, L. Nealy, J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1997**, *387*, 202-205. b) L. A. Dickinson, R. J. Gulizia, J. W. Trauger, E. E. Baird, D. E. Mosier, J. M. Gottesfeld, P. B. Dervan *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895.

[3] C. C. C. Wang, U. Ellervik, P. B. Dervan, *Bioorg. Med. Chem.* **2001**, *3*, 653-657.

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

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

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

[7] a) B. L. Iverson, P. B. Dervan, *Nucl. Acids Res.* **1987**, *15*, 7823-7830. b) A. M. Maxam, W.S. Gilbert, *Methods Enzymol.* **1980**, *65*, 499-560.

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

Chapter 6

Recognition of A•T Sequences by Hairpin Polyamides with Pyrrole / Hydroxypyrrole Pairs at the Terminal Position

Abstract

Recently ring pairs have been developed to target T•A and C•G in addition to the original G•C base pairs at the terminal positions of hairpin polyamides. In order to evaluate hairpin polyamides capable of distinguishing A•T base pairs at the terminal position, four hairpin polyamides, PyPyPyPy-(R)^{H_2N}γ-ImPyPyHp-β-Dp (**1**), PyPyPyPy-(R)^{H_2N}γ-ImPy-β-Py-β-Dp (**2**), BzPyPyPy-(R)^{H_2N}γ-ImPyPyHp-β-Dp (**3**), and BzPyPyPy-(R)^{H_2N}γ-ImPy-β-Py-β-Dp (**4**), were synthesized. The equilibrium association constants (K_a) for these compounds are assessed and the all rings polyamides **1** and **3** showed no preference for the target site 5'-T**A**TACA-3' over 5'-T**T**TACA-3' although ten-fold preference is shown over 5'-T**G**TACA-3' and 5'-T**C**TACA-3'. The addition of a flexible residue in polyamides **2** and **4** did not improve the targeting of the intended match site 5'-T**A**TACA-3'. Instead the β-alanine containing polyamides showed five to fifty-fold lower affinity for all four sites compared to the parent compounds.

Introduction

The ability to distinguish T•A from A•T base pairs by synthetic ligands which bind in the minor groove of DNA is an essential milestone for targeting specific A/T rich sequences, within promoters for gene regulation studies.[1] We recently reported that the 2-hydroxypyrrole/pyrrole pair (Hp/Py) for distinguishing T•A from A•T is most suitable for internal positions in polyamides.[2] The Hp ring at the N-terminal position affords lower affinity and sequence specificity for A•T/T•A vs G•C/C•G than anticipated. T•A recognition at the terminal position of hairpin polyamides was finally achieved by substituted 2-hydroxy-benzamide/pyrrole pairs.[3] To target C•G base pairs at the N-terminal position, we recently paired a pyrrole ring with an imidazole ring.[4] To our surprise the polyamide bound its target site 5'-T**C**TACA-3' with little specificity versus its single base pair mismatch 5'-T**G**TACA-3'. We found that a flexible β-alanine was necessary to bring the key imidazole ring back to register in order to differentiate 5'-T**C**TACA-3' from its mismatch 5'-T**G**TACA-3'.

With the successful development of ring pairs to target T•A and C•G in addition to the original G•C at the terminal position, we turn our attention to designing hairpin polyamides capable of recognizing A•T base pairs using pyrrole/hydroxypyrrole pairs. We describe here a series of four hairpin polyamides with hydroxypyrrole in the C-terminal position. The hydroxypyrrole is paired with an N-terminal pyrrole or benzoyl residue. Two compounds, with a flexible β-alanine next to the hydroxypyrrole, were also included in this study with the expectation that the hydroxypyrrole ring might be out of register as in the case of compounds designed to target C•G (Figure 6.1, 6.2).

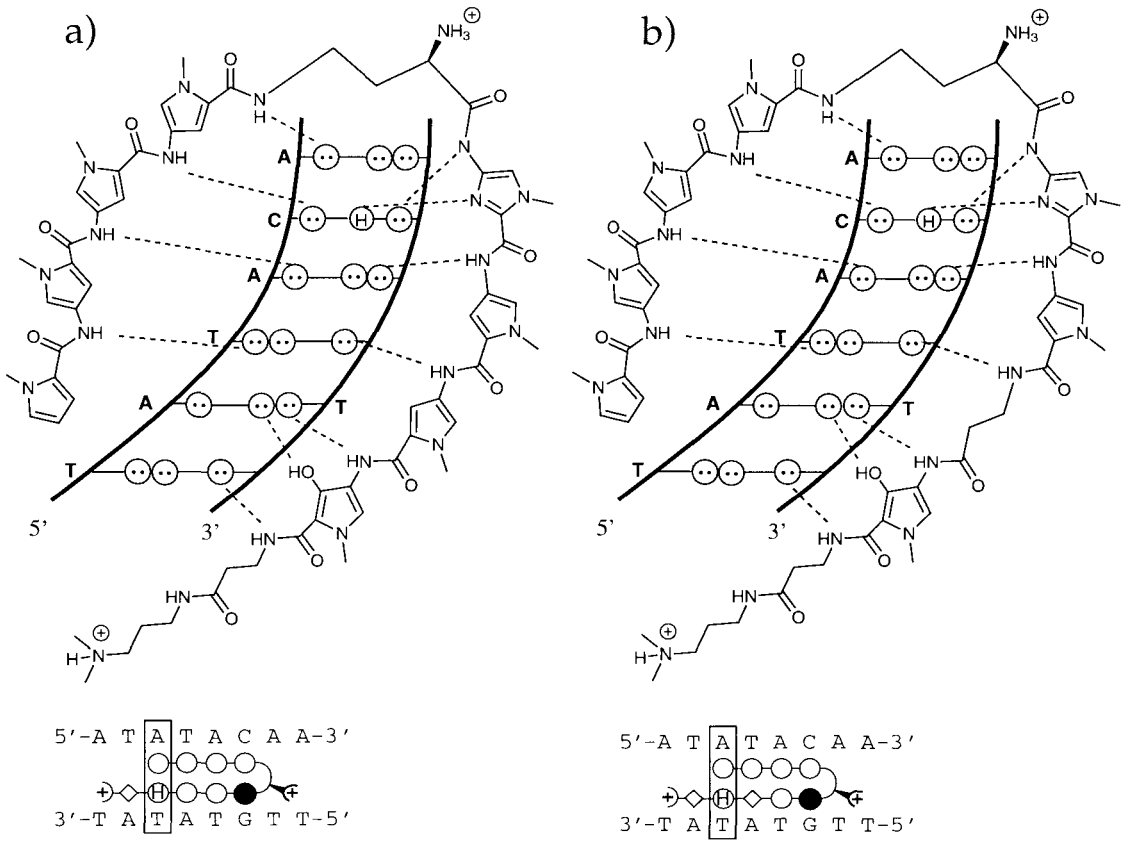


Figure 6.1. a) Hydrogen bonding model of the polyamide:DNA complex between the eight-ring hairpin polyamide PyPyPyPy-(R)H_2Nγ-ImPyPyHp-β-Dp (**1**) with the 5'-TATACA-3'-site. **b)** Hydrogen bonding model of the polyamide:DNA complex between the eight-ring hairpin polyamide PyPyPyPy-(R)H_2Nγ-ImPy-β-Hp-β-Dp (**2**) with the 5'-TATACA-3'-site. A circle with two dots represents the lone pairs of N3 of purines and O2 of cytosine. Two touching circles with dots represent the two lone pairs of the O2 of thymine. Circles containing an H represent the exocyclic amino hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

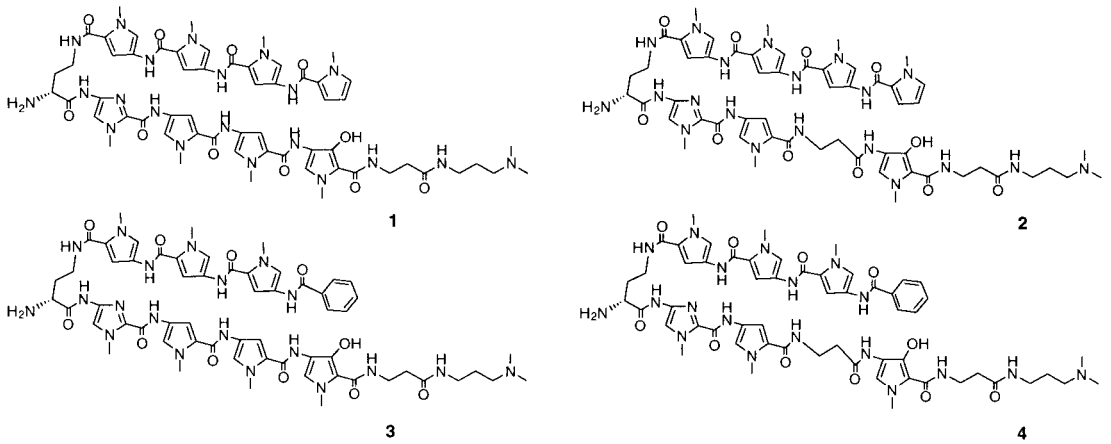


Figure 6.2. Structures of polyamides. PyPyPyPy-(R)H_2Nγ-ImPyPyHp-β-Dp (**1**), PyPyPyPy-(R)H_2Nγ-ImPy-β-Py-β-Dp (**2**), BzPyPyPy-(R)H_2Nγ-ImPyPyHp-β-Dp (**3**), BzPyPyPy-(R)H_2Nγ-ImPy-β-Py-β-Dp (**4**),

The plasmid pCW15 was designed to contain four six-base pair recognition sites 5'-TNTACA-3' (where N=T, A, G, C) which differ at a single common position allowing for comparison of the affinities between different terminal pairs and the four Watson Crick base pairs in the minor groove of DNA (Figure 6.3).

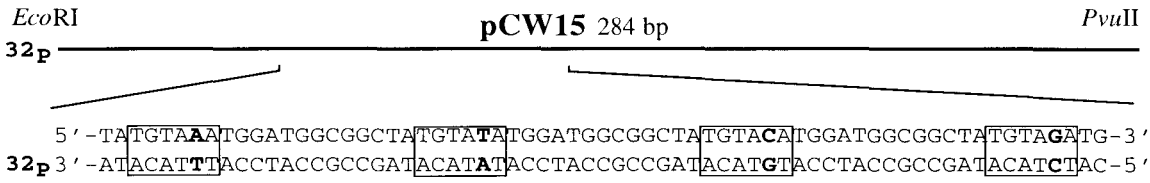


Figure 6.3. 284 base pair *Eco*RI/*Pvu*II restriction fragment derived from plasmid pCW15. The targeted six-base pair recognition sites are shown in boxes.

Results and Discussions

Polyamide synthesis. BocHN-PyPyPy-(R)^{Fmoc}γ-ImPyPyOp-β-PAM-resin and BocHN-PyPyPy-(R)^{Fmoc}γ-ImPy-β-Op-β-PAM-resin were synthesized in a stepwise manner from Boc-β-alanine-PAM resin (0.55 mmol/g) using solid-phase methodology[4] in 16 steps (Figure 4). Hydroxypyrrole amino acid residues were introduced as *orthogonally*

protected 3-methoxy-1-methylpyrrole (Op) derivatives.[6] Terminal residues were coupled to a sample of the resin which was then cleaved by aminolysis with ((dimethyl)amino)propylamine (55°C, 16 hours) and purified by reversed-phase HPLC. The methyl-protected polyamides were subsequently deprotected by treatment with BBr_3•SMe_2 in 1,2-dichloroethane[7] and purified by reversed-phase HPLC to afford PyPyPyPy-(R)H_2Nγ-ImPyPyHp-β-Dp (**1**), PyPyPyPy-(R)H_2Nγ-ImPy-β-Hp-β-Dp (**2**), BzPyPyPy-(R)H_2Nγ-ImPyPyHp-β-Dp (**3**) and BzPyPyPy-(R)H_2Nγ-ImPy-β-Hp-β-Dp (**4**) (Figure 6.2).

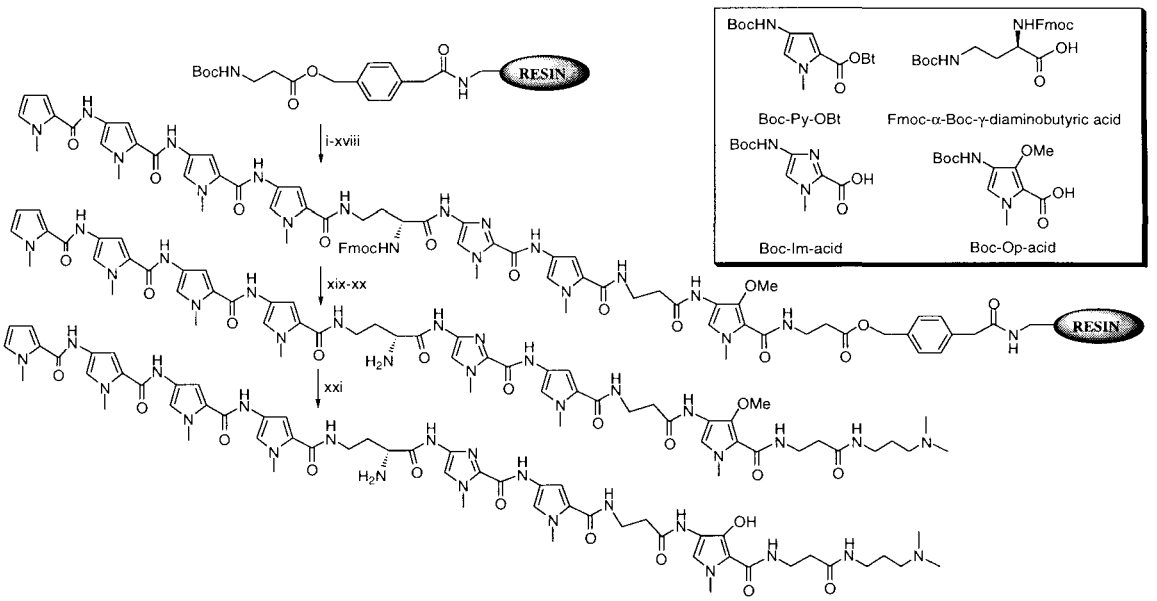


Figure 6.4. Solid-phase synthetic scheme for PyPyPyPy-(R)H_2Nγ-ImPy-β-Py-β-Dp (**2**), starting from commercially available Boc-β-PAM-resin: (i) 80% TFA/DCM, 0.4 M PhSH. (ii) Boc-Op-acid, HBTU, DIEA, DMF. (iii) 80% TFA/DCM, 0.4 M PhSH. (iv) Boc-β-Ala, HBTU, DIEA, DMF. (v) 80% TFA/DCM, 0.4 M PhSH. (vi) Boc-Py-OBt, DIEA, DMF. (vii) 80% TFA/DCM, 0.4 M PhSH. (viii) Boc-Im-acid, DCC, HOBt, DIEA, DMF. (ix) 80% TFA/DCM, 0.4 M PhSH. (x) Fmoc-α-Boc-γ-diaminobutyric acid, HBTU, 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-Py-OBt, DIEA, DMF. (xvii) 80% TFA/DCM, 0.4 M PhSH. (xviii) Py-acid, HBTU, DIEA, DMF. (xix) Piperidine:DMF 3:1. (xx) (Dimethylamino)propylamine, 55°C, 16 hours.(xxi) BBr_3•SMe_2, 1,2-dichloroethane.

Quantitative DNase I footprinting titrations. Quantitative DNase I footprint titrations were performed on the 3'-^{32}P-end-labeled 284 base pair *Eco*RI/*Pvu*II restriction fragment of the pCW15 plasmid[2,3] to determine the equilibrium association constant (K_a) of each hairpin polyamide to the four different binding sites 5'-TNTACA-3' (where N=T, A, G, C). Both compounds **1** and **3** did not differentiate the target match site 5'-T**A**TACA-3' from the single base pair mismatch site 5'-T**T**TACA-3' although the two other single base pair mismatch sites 5'-T**G**TACA-3' and 5'-T**C**TACA-3' were bound with ten-fold lower affinity. The positioning of the hydroxypyrrole ring at the C-terminus of the hairpin could possibly put it out of register with the DNA. Therefore, compounds **2** and **4** with flexible β-alanine residue were also tested. Footprinting experiments showed that the β-alanine containing polyamides **2** and **4** did not target 5'-T**A**TACA-3' over their mismatch sites. Moreover, the overall affinities for all four sites are five to fifty-fold lower for the β-alanine containing polyamides compared to their parent compounds.

Table 6.1. Equilibration Association Constants (M^{-1})[a]

Polyamide	5'-T**T**TACA-3'	5'-T**A**TACA-3'	5'-T**G**TACA-3'	5'-T**C**TACA-3'
Py/Hp (1)	2.8×10^9	1.1×10^9	1.1×10^8	1.0×10^8
Py/Hp with beta (2)	1.0×10^8	7.6×10^7	4.7×10^7	1.0×10^7
Bz/Hp (3)	4.3×10^9	2.4×10^9	4.0×10^8	3.2×10^8
Bz/Hp with beta (4)	8.5×10^7	4.2×10^7	4.2×10^7	3.7×10^7

[a]Values reported are mean values from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. The assays were perfomed at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

The lack of hairpin polyamides capable of binding 5'-T**A**TACA-3' tighter than its three mismatches suggests the need for the design of alternative ring pairs. The attempt to use flexible β-alanine residue in place of a pyrrole ring resulted in a poorer binding polyamide. This result suggests that β -alanine is incompatible with hydroxypyrrole rings.

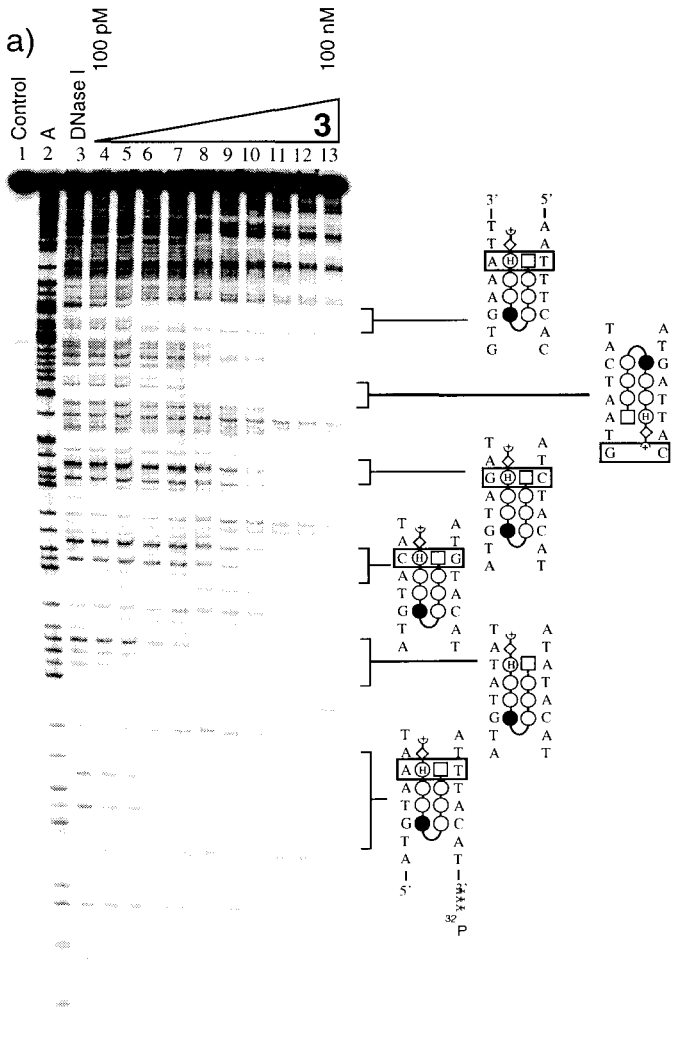


Figure 6.5. (a) Quantitative DNase I footprint titration experiment with BzPyPyPy-(R)H_2Nγ-PyPyPyPyHp-β-Dp (**3**) on the 284 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pCW15: lane 1, intact DNA; lane 2, A specific reaction; lane 3, DNase I standard; lanes 4-13, 100pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM BzPyPyPy-(R)H_2Nγ-PyPyPyPyHp-β-Dp (**3**) (b) Binding isotherms for polyamide **3.** The four sites 5'-T**T**TACA-3', 5'-T**A**TACA-3', 5'-T**G**TACA-3' and 5'-T**C**TACA-3' sites that were analyzed are shown on the right side of the gel.

Experimental Section

Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-PAM-Resin), N,N'-Dicyclohexylcarbodiimide (DCC), N-Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and Boc-β-alanine were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA) and *N,N*-dimethylformamide (DMF) were purchased from Applied Biosystems. The DMF was distilled under reduced pressure prior to synthesis. (R)-2-Fmoc-4-Boc-diaminobutyric acid was from Bachem, dichloromethane (DCM) was reagent grade from EM, thiophenol (PhSH) and dimethylaminopropylamine from Aldrich, and trifluoroacetic acid (TFA) from Halocarbon. ^{1}H NMR were recorded on a Varian Mercury 300 instrument. Chemical shifts are reported in ppm downfield from the signal for Me_4Si, with reference 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 on a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5μm, 300 × 4.6 mm reversed-phase column in 0.1% (w/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 × 100 mm, 100μm C_{18} column, 0.1% (w/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. Reagent-grade chemicals were used unless otherwise stated.

BocHN-PyPyPy-(R)$^{Fmoc\gamma}$-ImPyPyOp-β-PAM-resin and BocHN-PyPyPy-(R)$^{Fmoc\gamma}$-ImPy-β-Op-β-PAM-resin. BocHN-PyPyPy-(R)$^{Fmoc\gamma}$-ImPyPyOp-β-PAM-resin (0.33 mmol/g) and BocHN-PyPyPy-(R)$^{Fmoc\gamma}$-ImPy-β-Op-β-PAM-resin (0.33 mmol/g) were

synthesized in a stepwise fashion from 0.55 mmol/g Boc-β-PAM-resin by manual solid-phase methods.[4]

Procedure for cleavage from the resin. After the coupling was completed the resin was filtered off the reaction and washed with DMF (2 × 30 seconds). DMF (1 mL) and piperidine (3 mL) were added and the mixture shaken for 30 minutes at 22°C. The resin was filtered off and washed with DMF (2 × 30 seconds), DCM (3 × 30 seconds), MeOH (1 × 30 seconds), Et_2O (1 × 30 seconds) and dried *in vacuo*. The resin was treated with (dimethylamino)propylamine (1 mL) with periodic agitation at 55°C for 16 hours. The reaction mixture was then filtered to remove the resin, TFA (0.1% (w/v), 7 mL) added and the resulting solution purified by reversed-phase HPLC. The pure compounds were recovered as white powders upon lyophilization of the appropriate fractions.

Deprotection of methyl-protected polyamides. The polyamide was dissolved in 1,2-dichloromethane (1 mL) and BBr_3•SMe_2 (20 mg) was added and the mixture was heated to 80°C for 3 hours.[5] Water (3 mL) was added and all volatiles were removed (high vacuum, 40°C). The residue was dissolved in 1 mL CH_3CN and TFA (7 mL, 0.1% (w/v)) and purified by reverse phase HPLC.

PyPyPyPy-(R)^{H_2N}γ-ImPyPyHp-β-Dp (1). 1-methylpyrrole-2-carboxylic acid (50 mg, 0.40 mmol) was dissolved in DMF (1 mL) and HBTU (75 mg, 0.20 mmol) and DIEA (0.5 mL) were added and the mixture stirred for 10 minutes and added to deprotected BocHN-PyPyPy-(R)^{Fmoc}γ-ImPyPyOp-β-PAM-resin (100 mg) and the mixture shaken for 4 hours at 22°C. The compound was then cleaved from the resin according to the general procedure. Preparative HPLC gave PyPyPyPy-(R)^{H_2N}γ-ImPyPyOp-β-Dp (26 mg, 62% isolated yield). UV λ_{max} (H_2O) 240, 312 (69500); MALDI-TOF-MS calcd. for $C_{60}H_{75}N_{21}O_{11}$ (M+H): 1266.6 found: 1266.8. The methyl-protected PyPyPyPy-(R)^{H_2N}γ-ImPyPyOp-β-Dp (3.8 mg, 3.0 μmol) was deprotected according to the general procedure to

give **1** (1.4 mg, 37% isolated yield). UV λ_{max} (H_2O) 312 (69500); MALDI-TOF-MS calcd. for $C_{59}H_{73}N_{21}O_{11}$ (M+H): 1252.6; found: 1252.4.

PyPyPyPy-(R)^{H_2N}γ-ImPy-β-Hp-β-Dp (2). 1-methylpyrrole-2-carboxylic acid (50 mg, 0.40 mmol) was dissolved in DMF (1 mL) and HBTU (75 mg, 0.20 mmol) and DIEA (0.5 mL) were added and the mixture stirred for 10 minutes and added to deprotected BocHN-PyPyPy-(R)^{Fmoc}γ-ImPy-β-Op-β-PAM-resin (100 mg) and the mixture shaken for 1.5 hours at 37°C. The compound was then cleaved from the resin according to the general procedure. Preparative HPLC gave PyPyPyPy-(R)^{H_2N}γ-ImPy-β-Op-β-Dp (14 mg, 36% isolated yield). UV λ_{max} (H_2O) 240, 312 (60830); MALDI-TOF-MS calcd. for $C_{57}H_{74}N_{20}O_{11}$ (M+H): 1215.6 found: 1215.6. The methyl-protected PyPyPyPy-(R)^{H_2N}γ-ImPy-β-Op-β-Dp (3.8 mg, 3.1 μmol) was deprotected according to the general procedure to give **2** (1.4 mg, 38% isolated yield). UV λ_{max} (H_2O) 242, 312 (60830); MALDI-TOF-MS calcd. for $C_{56}H_{72}N_{20}O_{11}$ (M+H): 1201.6; found: 1201.54.

BzPyPyPy-(R)^{H_2N}γ-ImPyPyHp-β-Dp (3). Benzoic acid (70 mg, 0.57 mmol) was dissolved in DMF (1.0 mL) and HBTU (75 mg, 0.20 mmol) and DIEA (0.5 mL) were added. The mixture was allowed to stir for 10 minutes and added to the BocHN-PyPyPy-(R)^{Fmoc}γ-ImPyPyOp-β-PAM-resin (100 mg). The mixture was shaken for 5.0 hours at 22°C, and the polyamide was cleaved from the resin according to the general procedure. Preparative HPLC gave BzPyPyPy-(R)^{H_2N}γ-ImPyPyOp-β-Dp (16 mg, 39% isolated yield). UV λ_{max} (H_2O) 240, 312 (69500); MALDI-TOF-MS calcd. for $C_{61}H_{74}N_{20}O_{11}$ (M+H): 1263.6 found: 1263.7. The methyl-protected BzPyPyPy-(R)^{H_2N}γ-ImPyPyOp-β-Dp (6.4 mg, 5.1 μmol) was deprotected according to the general procedure to give **3** (1.4 mg, 19% isolated yield). UV λ_{max} (H_2O) 312 (69500); MALDI-TOF-MS calcd. for $C_{60}H_{72}N_{20}O_{11}$ (M+H): 1249.57; found: 1249.7.

BzPyPyPy-(R)^{H_2N}γ-ImPy-β-Hp-β-Dp (4). A mixture of DMF (1 mL), DIEA (0.5 mL) and benzoyl chloride (0.05 mL, 0.4 mmol) was added to deprotected Boc-PyPyPy-(R)^{Fmoc}γ-ImPyβIm-β-PAM-resin (100 mg) and shaken for 20 minutes at 22°C and the polyamide was cleaved from the resin according to the general procedure. Preparative HPLC gave BzPyPyPy-(R)^{H_2N}γ-ImPy-β-Op-β-Dp (16 mg, 41% isolated yield). UV λ_{max} (H_2O) 240, 312 (60830); MALDI-TOF-MS calcd. for $C_{58}H_{73}N_{19}O_{11}$ (M+H): 1212.6 found: 1212.7. The methyl-protected BzPyPyPy-(R)^{H_2N}γ-ImPy-β-Op-β-Dp (2.7 mg, 2.1 μmol) was deprotected according to the general procedure to give **4** (1.5 mg, 59% isolated yield). UV λ_{max} (H_2O) 242, 312 (68030); MALDI-TOF-MS calcd. for $C_{57}H_{71}N_{19}O_{11}$ (M+H): 1198.6; found: 1198.4.

Acknowledgement

We are grateful to the National Institutes of Health (GM-27681) for research support and a research traineeship award (GM-08501) to C.C.C.W. and the Knut and Alice Wallenberg Foundation for a post-doctoral fellowship to U.E. We thank G. M. Hathaway for MALDI-TOF mass spectrometry.

References

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

2) Chapter 2.

3) Chapter 3.

4) Wang, C. C. C.; Ellervik, U.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, 3, 653-657.

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

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

7) Williard, P.G.; Fryhle, C.B., *Tetrahedron Lett*, **1980**, *21*, 3731-3734

Chapter 7

Sequence Specific Trapping of Topoisomerase I by DNA Binding Polyamide-Camptothecin Conjugates

The text of this chapter was taken in part from a publication coauthored with Prof. Peter B. Dervan.
(Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2001** in press.)

Abstract

Hairpin pyrrole-imidazole polyamides are synthetic ligands that bind in the minor groove of DNA with affinities and specificities comparable to DNA binding proteins. Three polyamide-camptothecin conjugates **1-3** with linkers varying in length between 7, 13 and 18 atoms were synthesized to trap the enzyme Topoisomerase I and induce cleavage at predetermined DNA sites. One of these, polyamide-camptothecin conjugate **3** at nanomolar concentration (50 nM) in the presence of Topo I (37°C) induces DNA cleavage between three and four base pairs from the polyamide binding site in high yield (77%).

Introduction

Pyrrole-imidazole polyamides that target DNA sequences in the promoter have been shown to inhibit transcription of specific genes.[1-3] The question arises whether sequence specific DNA binding polyamides could be designed to recruit endogenous cellular enzymes to modify DNA at specific gene sites reminiscent of RNase H cleavage of DNA•RNA hybrids in the antisense field.[4] Topoisomerase I (Topo I), an endogenous eukaryotic enzyme essential in replication and transcription, cleaves and religates the phosphodiester backbone of DNA in an ATP independent fashion.[5-7] The catalytic activity of Topo I involves a three step process: (i) cleavage of one strand of DNA where upon the enzyme is covalently linked to the 3' phosphoryl end of the nick through a tyrosine residue termed the cleavable complex, (ii) passage of a segment of DNA through the nick, and (iii) resealing of the DNA break. This sequence is interrupted by the presence of the natural product camptothecin (Cpt) which stabilizes the cleavable complex and prevents religation of the DNA.[8] Camptothecin-stabilized Topo I –DNA adducts have been shown to cause premature termination of transcription elongation by RNA polymerase II.[9] A polyamide-camptothecin conjugate bound site-specifically in the minor groove within a "coding region" of a specific gene might trap a covalent adduct of the Topo I enzyme with DNA. The site specific protein-DNA adduct could be a gene specific road block for the elongation of the RNA polymerase. Targeting Topoisomerase I to DNA by triple helix forming oligonucleotide conjugates have been previously reported by Matteucci[10] and Helene.[11] These oligonucleotide-camptothecin (and rebeccamycin) conjugates bind in the major groove of DNA. It would be interesting to compare major vs minor grove positioning of the Topo I poison. We report here our efforts to characterize the properties of a minor groove binding hairpin polyamide-camptothecin conjugate with regards to the structural requirements in the linker region for sequence specifically trapping the Topo I mediated cleavable complex.[12]

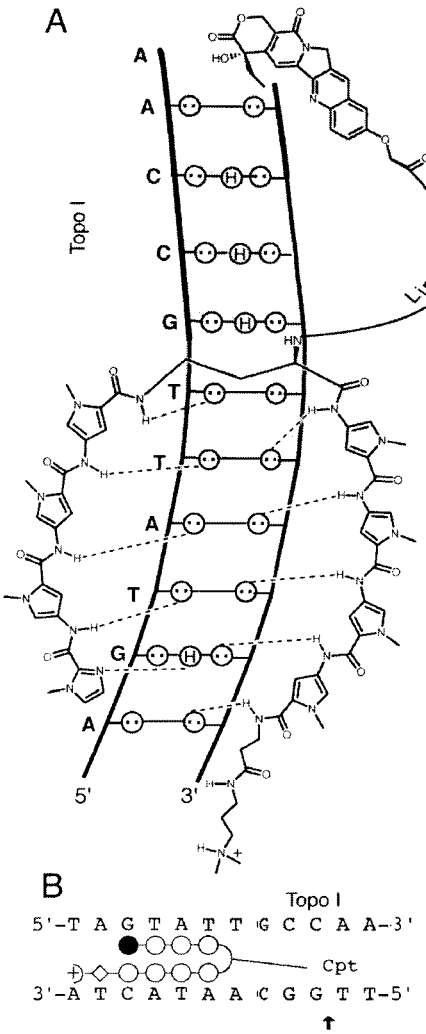


Figure 7.1. (A) Hydrogen-bonding model of a polyamide-camptothecin conjugate with a generic linker bound to the minor groove of 5'-AGTATT-3' and Topo I. Circles with two dots represent the lone pairs of N3 purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. (B) Binding model of a polyamide-camptothecin conjugate. Shaded and nonshaded circles denote imidazole (Im) and pyrrole (Py) rings, respectively. Diamonds, small rectangle and big oval represent β-alanine (β), camptothecin (Cpt), and Topoisomerase I (Topo I) respectively. (R)-2,4-diaminobutyric acid (γ) and dimethylaminopropylamine (Dp) are depicted as a curved line and a plus sign, respectively.

Currently there are no three-dimensional structures of camptothecin bound to the DNA-Topo I cleavable complex although several groups have presented models.[6,13] Since the requirements for camptothecin presentation are unclear, we designed and synthesized three polyamide conjugates **1-3** with linkers varying in length between 7, 13 and 18 atoms and chemical composition (alkyl chain vs hydroxy alkyl chain). For synthetic ease, camptothecin was attached to the hairpin turn. The eight ring pyrrole-imidazole polyamide, ImPyPyPy-γ-PyPyPyPy-β-Dp (where Im is N-methylimidazole, Py is N-methylpyrrole, γ is γ-aminobutyric acid, β is β-alanine, and Dp is dimethylaminopropylamine); binds its match site 5'-AGTATT-3' and its mismatch site 5'-AGTACT-3' with equilibrium association constants of 3.5 x 10^9 M^{-1} and 5.0 x 10^8 M^{-1}, respectively.[14] We chose to modify camptothecin acid at C10 which has been shown by structure activity studies not to affect the ability of camptothecin to act as a Topo I poison.[15,16]

A plasmid pCW1 was constructed for the cleavage assay with polyamide-camptothecin conjugates **1-3** with one six-base pair match site 5'-AGTATT-3', and two mismatch sites 5'-TGGAAA-3' (single) and 5'-TGGACA-3' (double) according to the pairing rules where Im/Py pairing binds G/C base pair and Py/Py are degenerate for A/T and T/A base pairs. The polyamide is oriented on the match site with the camptothecin to the 3' side of the top strand. The sequences flanking the polyamide binding sites are 5'-GCCAAGTG-3'. Studies have shown that camptothecin induced Topo I cleavage occurs preferentially but not exclusively between T and G.[17,18]

Results and Discussion

Synthesis of polyamide-camptothecin conjugate 1-3. The polyamide conjugates **1-3** were synthesized in a stepwise manner from Boc-β-alanine Pam resin using Boc-chemistry protocols (Figure 7.3).[19] After the completion of the polyamide synthesis, the resin was split into three portions and three amino acid straight chain

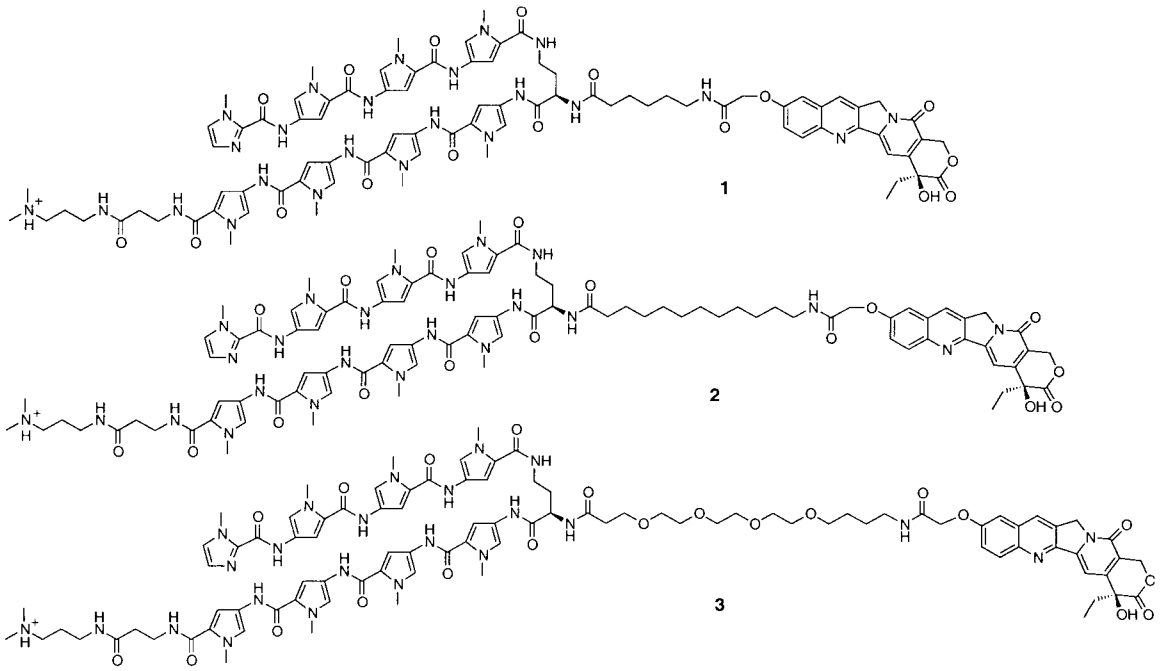


Figure 7.2. Structures of the hairpin polyamide-camptothecin conjugates. PA-CA5-Cpt (**1**), PA-CA11-Cpt (**2**), PA-EG4-Cpt (**3**).

linkers were incorporated. The three polyamide conjugates were each cleaved by a single step aminolysis reaction with dimethylaminopropylamine and subsequently purified by reverse phase HPLC chromatography. A camptothecin analogue **6** with a carboxy functionality suitable for conjugation was synthesized from 10-hydroxycamptothecin.[16a] Alkylation of 10-hydroxycamptothecin was accomplished with ethyl bromoacetate in refluxing acetone in the presence of anhydrous K_2CO_3 in 60% yield and the ester intermediate was hydrolyzed to form carboxymethyloxycamptothecin **6** in 90% yield. Activation of the camptothecin acid **6** was accomplished by dicyclohexylcarbodiimide (DCC) and N-hydroxysuccinimide (NHS). Polyamides with a primary amine on the turn was allowed to react with the activated acid in the presence of diisopropylethylamine (DIEA) afforded the polyamide-camptothecin conjugates **1-3**. The conjugates were

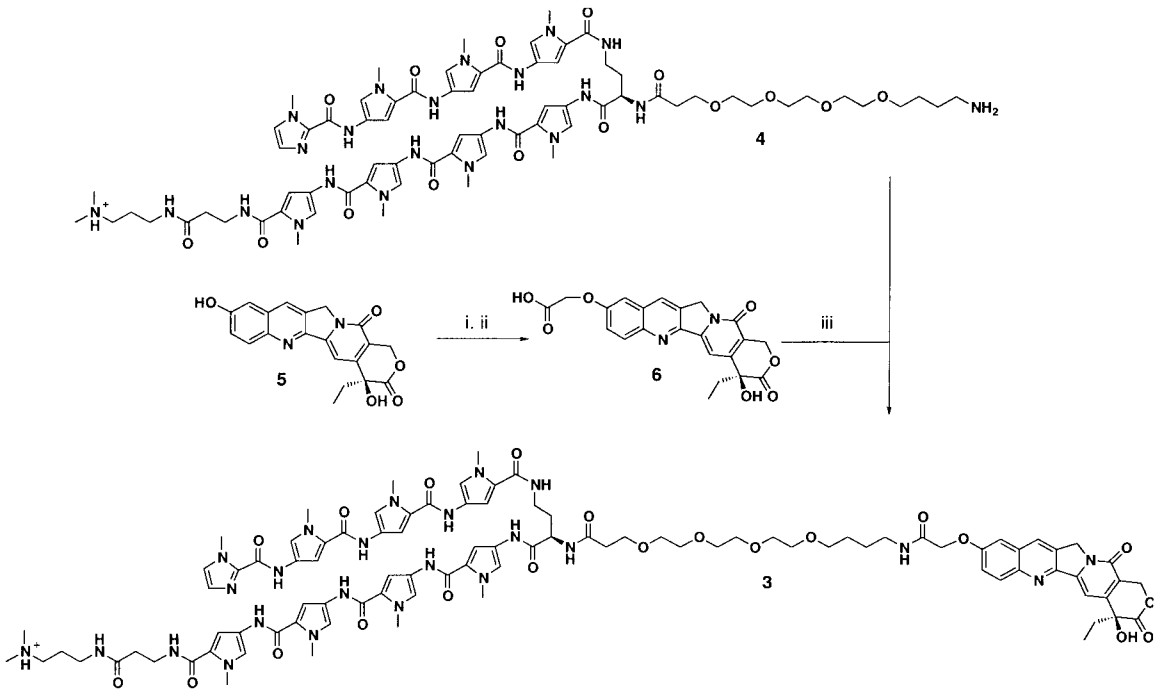
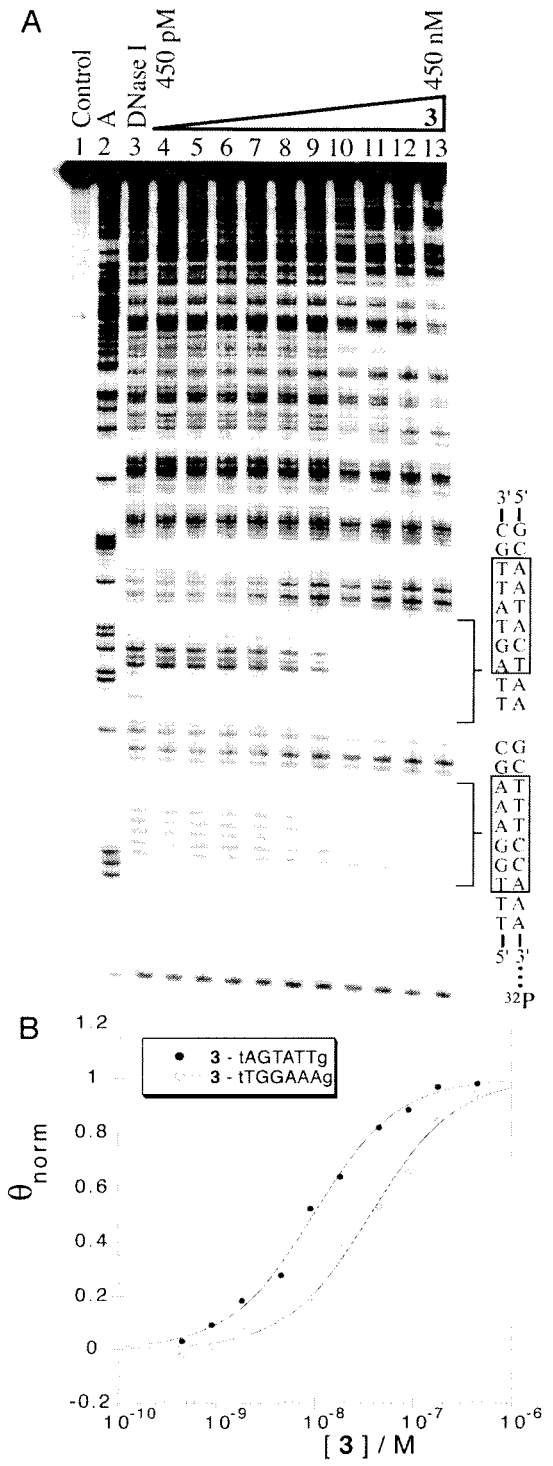


Figure 7.3. Synthetic scheme for polyamide-camptothecin conjugate **3** (i) ethylbromoacetate, K_2CO_3 60%; (ii) EtOH, H_2O, K_2CO_3, 90%; (iii) camptothecin acid **6**, DCC, HOBt 24% recovery.

purified by preparatory reverse phase HPLC. Characterization and identity was confirmed by MALDI-TOF-MS.

Determination of binding affinities by Quantitative DNase I footprinting titration experiments. Quantitative DNase I footprint titrations were performed on the 300 base pair *Eco*RI/*Pvu*II restriction fragment of pCW1 to determine the equilibrium association constants of the polyamide-camptothecin conjugates **1-3** for two sites 5'-AGTATT-3' and 5'-TGGAAA-3'(Table 1). The 300 base pair *Eco*RI/*Pvu*II restriction fragment includes the 115 base pair *Eco*RI/*Afl*II restriction fragment used for the Topo I cleavage experiment. For all three conjugates **1-3**, the match site 5'-AGTATT-3' was preferred over the single and double mismatch sites although we note the specificity of the conjugate overall appears compromised relative to the parent. The three conjugates show

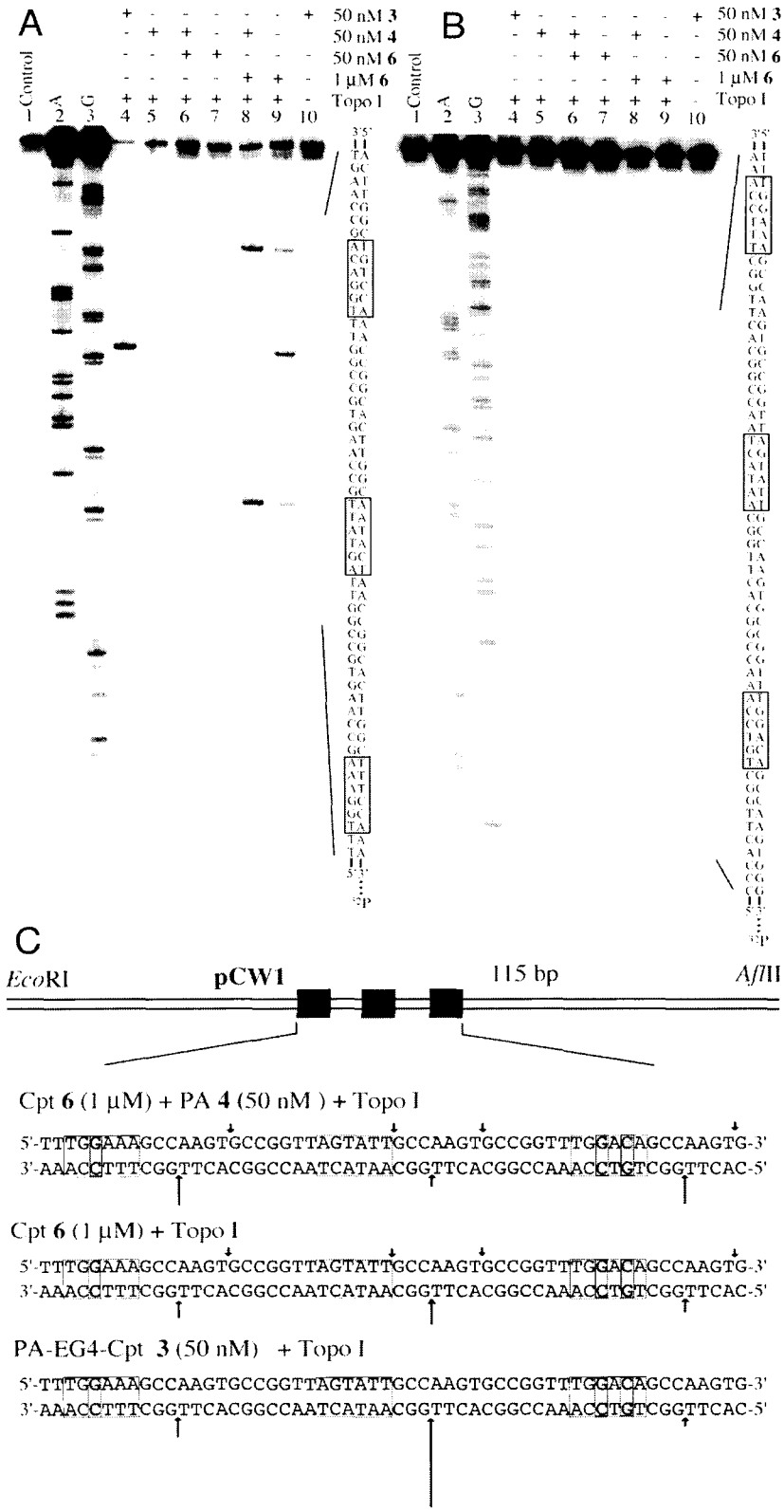
Figure 7.4. (A) Quantitative DNase I footprint titration experiment with PA-EG4-Cpt **3** on the 300 base pair *Eco*RI/*Pvu*II restriction fragment from plasmid pCW1: lane 1, intact DNA; lane 2, A specific reaction; lane 3, DNase I standard; lanes 4-13, 450 pM, 900 pM, 1.8 nM, 4.5 nM, 9 nM, 18 nM, 45 nM, 90 nM, 180 nM, 450 nM PA-EG4-Cpt **3**. The 5'-AGTATT-3' and 5'-TGGAAA-3' sites that were analyzed are shown on the right side of the gel. (B) Data from quantitative DNase I footprint titration experiments for PA-EG4-Cpt **3** binding to the two sites 5'-AGTATT-3' and 5'-TGGAAA-3'. θ_{norm} points were obtained using storage phosphor autoradiography and processed by standard methods.[14] The data for the binding of conjugate **3** to 5'- AGTATT -3' is indicated by filled circles and binding to 5'-TGGAAA-3' by open circles. The solid curves are best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm where $n = 1$ as previously described.[14]


A
Control
A
DNase I
450 pM
450 nM
3
1 2 3 4 5 6 7 8 9 10 11 12 13
3' 5'
C G
G C
T A
T A
A T
T A
G C
A T
T A
T A
C G
G C
A T
A T
A T
G C
G C
T A
T A
T A
5' 3'
32P
B
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
θnorm
3 - tAGTATTg
3 - tTGGAAAg
10-10
10-9
10-8
10-7
10-6
[3] / M

lower affinity for DNA than parent hairpin polyamide by a factor of 10 which is attributed to the attachment of the linker region at the γ turn.[20]

Polyamide-camptothecin conjugate recruits Topo I to DNA and cleaves DNA in high yield. The ability of the polyamide-camptothecin conjugates to recruit Topo I and cleave DNA sequence specifically was tested by a simple *in vitro* system.[10] A radio-labeled DNA restriction fragment generated from plasmid pCW1 was 3' labeled either on the top or bottom strand. The ligands were incubated with the DNA restriction fragment followed by addition of 10 units of calf thymus Topo I for 1 hour at 37°C. Following protein denaturation by treatment with SDS and protease workup, the DNA restriction fragments were analyzed by denaturing polyacrylamide gel electrophoresis. Analysis was first done on the bottom strand. Conjugate **3** at 50 nM concentration in the presence of 10 units of Topo I produces a distinct cleavage site between T^G located 3-4 base pairs to the 3' side of the polyamide binding site 5'-AGTATT-3' (Figure 7.5A, lane 4). Weaker cleavage proximal to the mismatch site 5'-TGGAAA-3' was also observed. Quantitation revealed a cleavage yield of 77% proximal to the match site 5'-AGTATT-3' and 16% proximal to the single base pair mismatch 5'-TGGAAA-3'. Remarkably only 7% of the DNA fragment is left intact. In control experiments with 50 nM polyamide **4** alone or 50 nM camptothecin acid **6** alone, no cleavage was observed. Indeed concentrations of 1 μM were needed to observe cleavage with camptothecin acid **6** and three sites were observed at the bottom strand (Figure 7.5A, lane 9). Moreover, addition of 50 nM polyamide **4** and 50 nM camptothecin **6** with Topo I does not induce a cleavage band. Therefore, a covalent linkage between the polyamide and camptothecin is required for efficient site specific cleavage. Finally incubation of the conjugate **3** without Topo I revealed no cleavage indicating that Topo I is necessary for the DNA cleavage.

Figure 7.5. Topoisomerase I cleavage assay with polyamide-camptothecin conjugate **3** on the 115 base pair *Eco*RI/*Afl*II restriction fragments from plasmid pCW1: (A) 3' ^{32}P end-labeled on the bottom strand. lane 1, intact DNA; lane 2, A specific reaction; lane 3, G specific reaction; lane 4, 50 nM of conjugate **3** incubated with DNA followed by Topo I treatment; lane 5, 50 nM polyamide **4** with Topo I; lane 6, 50 nM camptothecin acid **6** and 50 nM polyamide **4** with Topo I; lane 7, 50 nM camptothecin acid **6** with Topo I; lane 8, 1 μM camptothecin acid **6** and 50 nM polyamide **4** with Topo I; lane 9, 1 μM camptothecin acid **6** with Topo I; lane 10, 50 nM conjugate **3** incubated with DNA. (B) 3' ^{32}P end-labeled on the top strand. lanes 1-10, same as panel A except on different DNA. (C) Illustration of the 115 base pair restriction fragment with the position of the sequence indicated. Cleavage pattern for 1 μM camptothecin acid **6** and 50 nM polyamide **4** with Topo I; Cleavage pattern for 1 μM camptothecin acid **6** with Topo I; Cleavage pattern for 50 nM conjugate **3** (PA-EG4-CPT) with Topo I. Arrow length is proportional to cleavage yield. Boxes represent polyamide binding sites with grey boxes over mismatches.

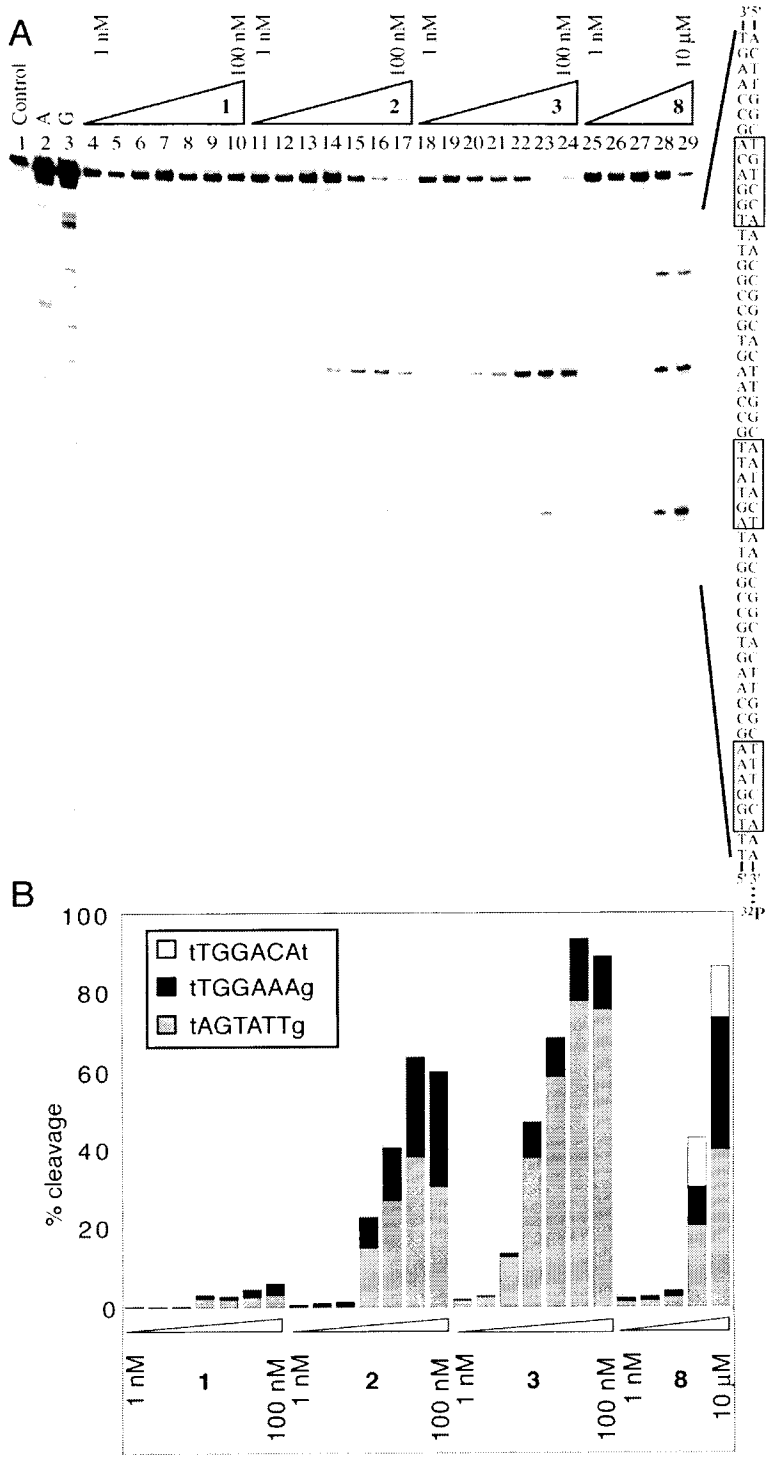

A
B
C
50 nM 3
50 nM 4
50 nM 6
1 μM 6
Topo I
Control
A
G
1 2 3 4 5 6 7 8 9 10
EcoRI
pCW1
115 bp
AflII
Cpt 6 (1 μM) + PA 4 (50 nM) + Topo I
5'-TTTGGAAAGCCAAGTGCCGGTTAGTATTGCCAAGTGCCGGTTTGGACAGCCAAGTG-3'
3'-AAACCTTTCGGTTCACGGCCAATCATAACGGTTCACGGCCAAACCTGTCGGTTCAC-5'
Cpt 6 (1 μM) + Topo I
5'-TTTGGAAAGCCAAGTGCCGGTTAGTATTGCCAAGTGCCGGTTTGGACAGCCAAGTG-3'
3'-AAACCTTTCGGTTCACGGCCAATCATAACGGTTCACGGCCAAACCTGTCGGTTCAC-5'
PA-EG4-Cpt 3 (50 nM) + Topo I
5'-TTTGGAAAGCCAAGTGCCGGTTAGTATTGCCAAGTGCCGGTTTGGACAGCCAAGTG-3'
3'-AAACCTTTCGGTTCACGGCCAATCATAACGGTTCACGGCCAAACCTGTCGGTTCAC-5'

Identical assays were then conducted on the top strand (Figure 7.5B). No cleavage was observed when 50 nM of conjugate **3** was incubated with 10 units of Topo I indicating that the cleavage occurs only in one strand (Figure 7.5B, lane 4). Camptothecin acid **6** at 1 μM created four distinct cleavage sites (Figure 7.5B, lane 9). However, the cleavage yield was very low as demonstrated by the large amount of intact DNA. The strand specific result leads us to conclude that the Topo I enzyme is oriented in one direction in the DNA/Topo I/polyamide-camptothecin complex. When 50 nM of polyamide **4** and 1 μM of camptothecin acid **6** were incubated together, one of the four cleavage sites disappeared (Figure 7.5B, lane 8). Analysis shows that the cleavage site is also the polyamide binding site suggesting that minor groove hairpin polyamide binding interferes sterically with the cleavage reaction.

Effects of the linker region on DNA cleavage yield. A series of polyamide-camptothecin conjugates differing only in the linker region were tested using the same *in vitro* assay (Figure 7.6). DNA cleavage yield decreased as the number of atoms in the linker region decreased: conjugate **3** (PA-EG4-Cpt) > conjugate **2** (PA-CA11-Cpt) > conjugate **1** (PA-CA5-Cpt). The trend is consistent within the ligand concentration range of 1 nM to 100 nM. For each conjugate, cleavage yield increased with increasing ligand concentration until leveling off at 50 nM. The longer and flexible linker in conjugate **3** possibly allows the camptothecin moiety to adopt a favorable orientation in the binding pocket and could be the reason for its higher activity.

Mapping Topo I cleavage site. Trapping of Topo I by the camptothecin polyamide conjugate is mapped on the DNA restriction fragment by comparing the reaction product to Maxam and Gilbert sequencing reaction taking into account that the

Figure 7.6. (A) Topoisomerase cleavage assay on conjugates **1-3** and camptothecin acid **6** on the *Eco*RI/*Afl*II restriction fragments from plasmid pCW1 labeled on the bottom strand: lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4-10, 1 nM, 2 nM, 5 nM, 10 nM , 20 nM, 50 nM, 100 nM of conjugate **1**; lane 11-17, 1 nM, 2 nM, 5 nM, 10 nM , 20 nM, 50 nM, 100 nM of conjugate **2**; lane 18-24, 1 nM, 2 nM, 5 nM, 10 nM , 20 nM, 50 nM, 100 nM of conjugate **3**; lane 25-29, 1 nM, 10 nM, 100 nM, 1 μM, 10 μM camptothecin acid **6,** all ligands were incubated with DNA and treated with Topo I.(B) Bar graph of cleavage efficiency by conjugates **1-3**: grey box represent % cleavage 3-4 base pairs from polyamide binding site 5'-AGTATT- 3', black box represent % cleavage 3-4 base pairs from 5'-TGGAAA- 3', white box represent % cleavage 4 base pairs from 5'-TGGACA- 3'.


A
Control
A
G
1 nM
100 nM
1 nM
100 nM
1 nM
100 nM
1 nM
10 µM
1
2
3
8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
32P
B
100
80
60
40
20
0
% cleavage
tTGGACAt
tTGGAAAg
tAGTATTg
1 nM
1
100 nM
1 nM
2
100 nM
1 nM
3
100 nM
1 nM
8
10 µM

Maxam and Gilbert reactions are gapping reactions while the Topo I reaction is an internucleotide cleavage reaction. Topo I cleavage creates a 5' OH group and a 3' Topo I linked fragment; therefore, the location of the cleavage site is accurately determined by labeling the DNA fragment at the 3' end using T4 polynucleotide kinase to add an extra phosphate to the 5' OH. The mapping of the cleavage site by the camptothecin polyamide conjugate shows the cleavage is at the 5'-ACTT^G-3' site which has been shown previously to be the consensus Topo I cleavage site (Figure 7.7).

Sequence effect of conjugate cleavage. Is the camptothecin polyamide conjugate sequence preference dictated only by the camptothecin unit? What would happen to regions where strong camptothecin Topo I sites occur without polyamide match sites nearby? To test this hypothesis a fragment of pUC19 was examined using a conjugate with a large 16 base pair binding site. The region of pUC19 examined has strong camptothecin sites but no binding site for the conjugate nearby. Even under micromolar concentrations no cleavage sites were observed, demonstrating the necessity of a polyamide site to create cleavage (Figure 7.8). Alternatively a region of the cancer related gene HER2 coding region was examined to ask the necessity of the camptothecin Topo I site adjacent to the polyamide site. Would it be sufficient to just have a polyamide binding site for the conjugate to trap Topo I and create cleavage? The experiment shows that the sequence composition of the flanking sequences has a strong effect to the cleavage yields and it is possible to observe conjugate cleavage at sites that are not strong Topo I sites (Figure 7.9). The two experiments show that neither the polyamide alone nor the camptothecin alone dictates the cleavage of the conjugate. To achieve maximum cleavage it is ideal to have both an optimal polyamide binding site and an optimal Topo I camptothecin site exist next to each other.

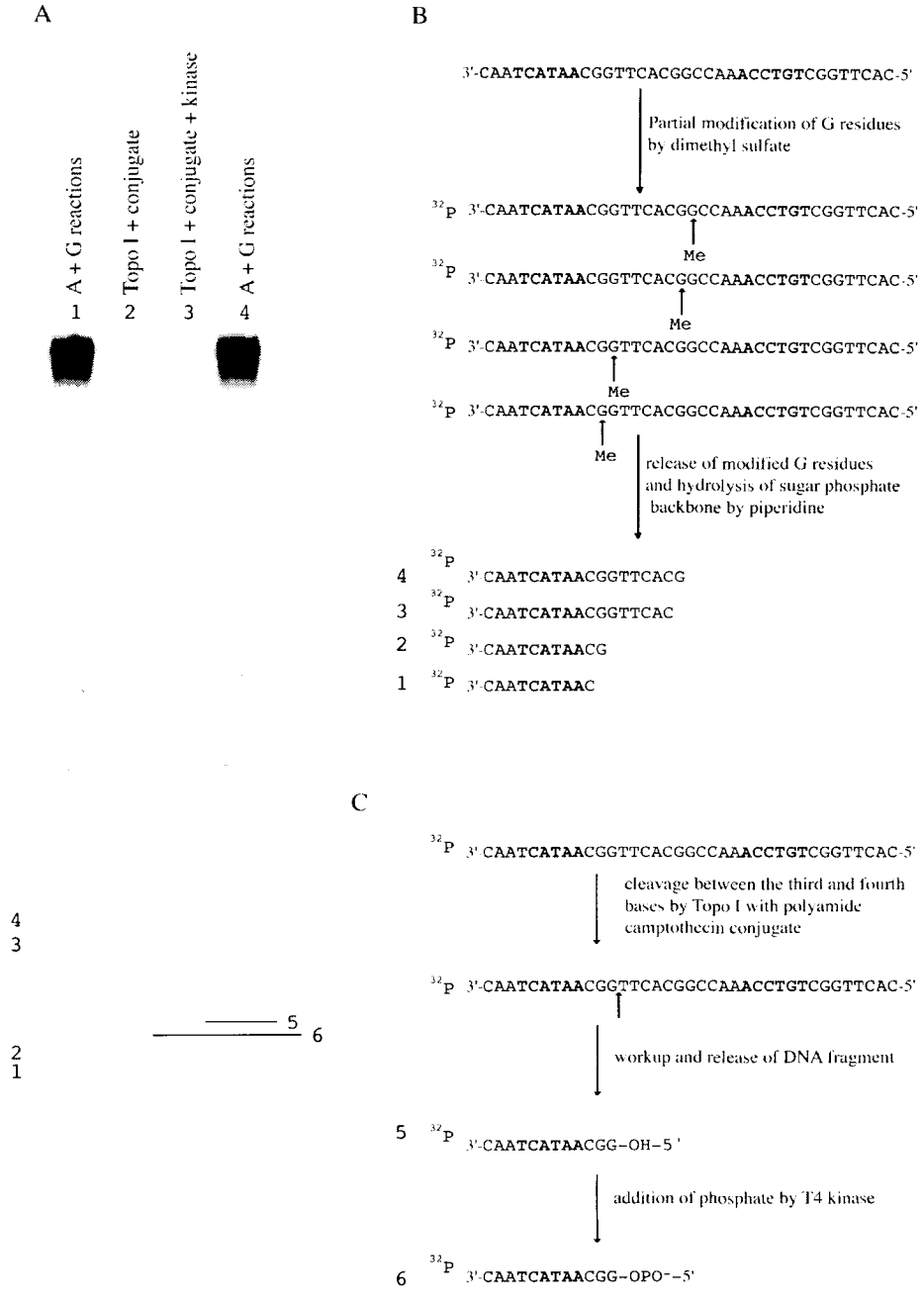
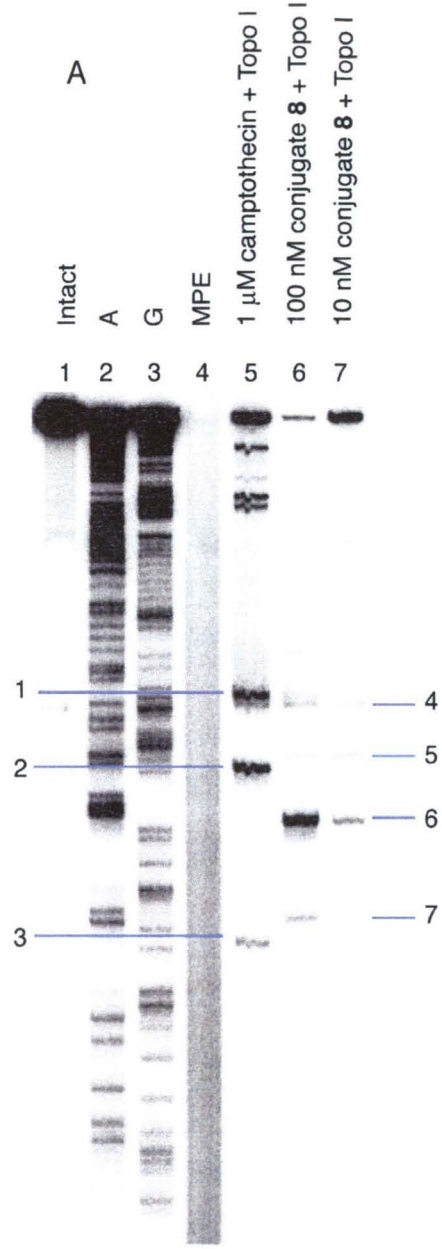


Figure 7.7. (A) Mapping of Topo I cleavage reaction. Lane 1, A + G reactions; lane 2, 10 nM of conjugate **3** with Topo I; lane 3, 10 nM of conjugate **3** with Topo I followed by reaction with T4 polynucleotide kinase; lane 4, A + G reactions. (B) Schematic of Maxam Gilbert G reaction. (C) Schematic of Topo I cleavage reaction.

Figure 7.8. (A) Topo I experiment with conjugate **7** on region of pUC19 containing no polyamide match site next to strong Topo I cleavage site. Lane 1, Intact; lane 2, A reaction; lane 3, G reaction; lane 4, Topo I; lane 5, 4 μM camptothecin **6** with Topo I; lane 6, 400 nM camptothecin **6** with Topo I; lane 7, 4 μM of conjugate **7** with Topo I; lane 8, 400 nM of conjugate **7** with Topo I. (B) Illustration of the restriction fragment. Cleavage position is illustrated by straight line.

A
Intact
A
G
Topo I
4 μM Cpt+ Topo I
400 nM Cpt + Topo I
4 μM Conjugate + Topo I
400 nM Conjugate + Topo I
1 2 3 4 5 6 7 8
B
7
C
5′-ACGCAATTAATGTGAGTTCTCACTCATTAGGCACCCCAGGCTTTACAC
3′-TGCGTTAATTACACTCAAGAGTGAGTAATCCGTGGGGTCCGAAATGTG
TTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAAC-3′
AAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTG-5′

Figure 7.9. (A) Topo I experiment with conjugate **8** with pHER1 containing conjugate match site flanked by different DNA sequences. The DNA restriction fragment is pieced together from different sections of the coding region of HER 2 containing polyamide match sites. Lane 1, Intact; lane 2 A reaction; lane 3, G reaction; lane 4, Topo I; lane 5, 1 μM camptothecin **6** with Topo I; lane 6, 100 nM of conjugate **8** with Topo I; lane 7, 10 nM of conjugate **8** with Topo I. (B) Illustration of conjugate **8** ImImPyPy-γ^{fpl}-ImPyPyPy-β-Dp. (C) Illustration of the restriction fragment. Cleavage position is illustrated by straight line and numbered. Bold sequence is the polyamide binding site.



C

5'-C**TGGACA**TGCTCCGCCG**AGGACA**ACTATGCCCGGAAC**AGGACA**CGATTTTGTA**AGGACA**TC
3'-G**ACCTGT**ACGAGGCGGC**TCCTGT**TGATACGGGCCTTG**TCCTGT**GCTAAAACAT**TCCTGT**AG
3 7 6 2

TTCCACAC**TGGTCA**CCTACAACAA**TGGTCT**GGGCATGGAG**TGGACT**GGCCCTCATTGCA-3'
AAGGTGTG**ACCAGT**GGATGTTGTT**ACCAGA**CCCGTACCTC**ACCTGA**CCGGGAGTAACGT-5'
5 4 1

Binding site size of camptothecin conjugate with Topo I. MPE•Fe(II) footprinting experiment carried out with camptothecin conjugate **3** and Topo I shows that the entire complex covers the 14 base pair site 5'-CACTT^GGCAATACT-3' (that is positions -5 to +9 with 0 being the position of the cleavage)(Figure 7.10, 7.11). X-ray structure of Topoisomerase I complex with DNA have been solved and the enzyme covers positions from -5 to +6 (Figure 7.12, 7.13). MPE shows conjugate **3** binds its six base pair site from +4 to +9. The footprint of the complex matches the sum of the footprint of the conjugate and footprint of Topo I seen in its crystal structure.

Implication for the design of Topo I hairpin polyamide conjugates. The results presented here reveal that polyamide-camptothecin conjugates can recruit Topo I to cleave single DNA sites in high yield. The difference in cleavage yields in conjugates **1-3** demonstrates the importance of linker length for conjugate activity. This sets the stage for the next step which is to test whether conjugates targeted to the coding region of specific genes can regulate transcription elongation which will be reported in due course.

Experimental Section

Materials. Fmoc-6-aminocaproic acid was purchased from Bachem, Fmoc-12-aminododecanoic acid was purchased from Aldrich and Boc 17-amino-4,7,10,13-tetraoxa-heptadecanoic acid was a generous gift of Dr. Anna Mapp.[21] 10-Hydroxycamptothecin was purchased from JS International. DNA restriction fragment labeling protocols, DNase I footprinting, determination of equilibrium association constants, plasmid construction and quantitation by storage phosphor autoradiography were as previously described.[14] Calf Thymus Topoisomerase I was obtained from Gibco BRL.

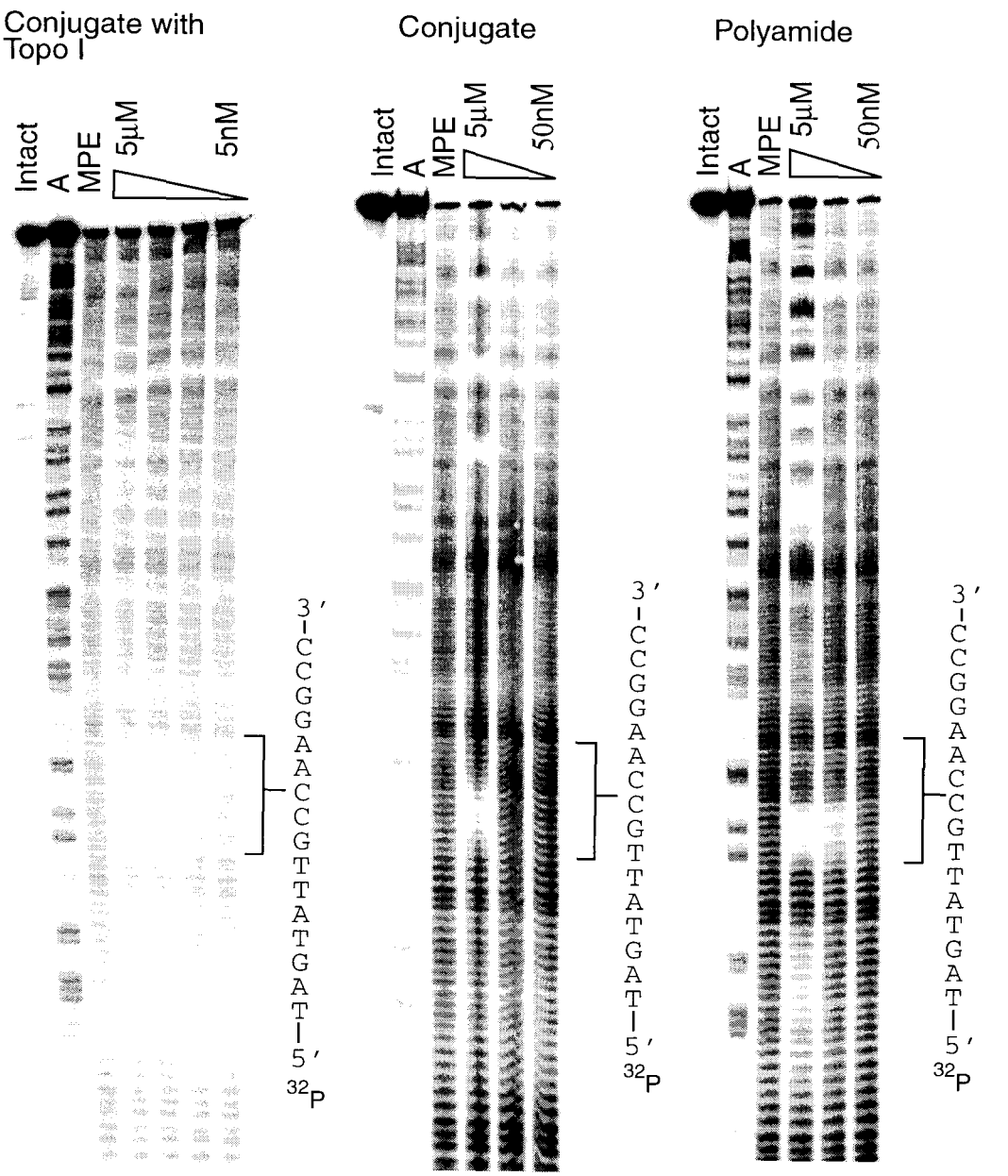
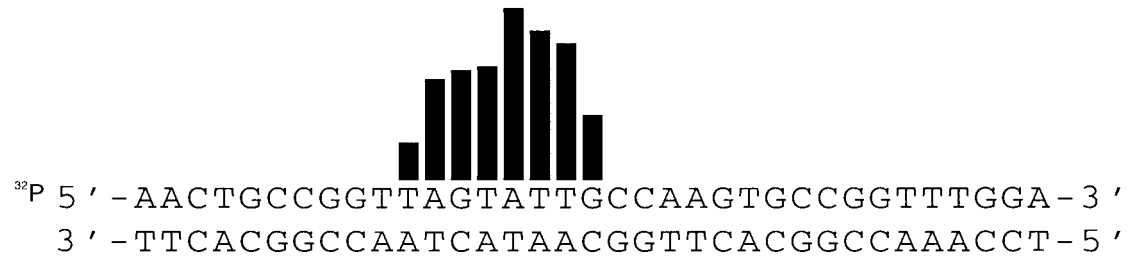


Figure 7.10. MPE•Fe(II) footprinting of conjugate **3** with Topo I, conjugate **3** alone, and polyamide **4** alone.

conjugate 3 + Topoisomerase I

conjugate 3 alone



polyamide 4 alone

Figure 7.11. Results of MPE•Fe(II) footprinting experiments. Bar heights are proportional to the extent of cleavage protection at the indicated base.

Figure 7.12. The structure of human Topo I in complex with DNA. The protein completely wraps around the DNA oligonucleotide contacting 4 base pair upstream and 6 base pair downstream of the cleavage site. Figure obtained from reference 6.

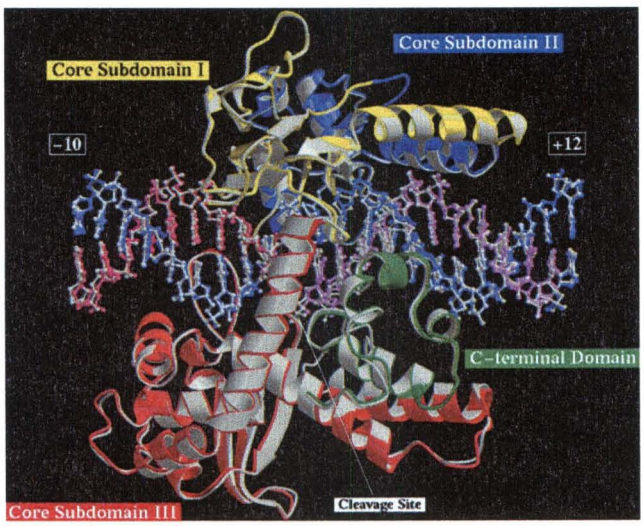
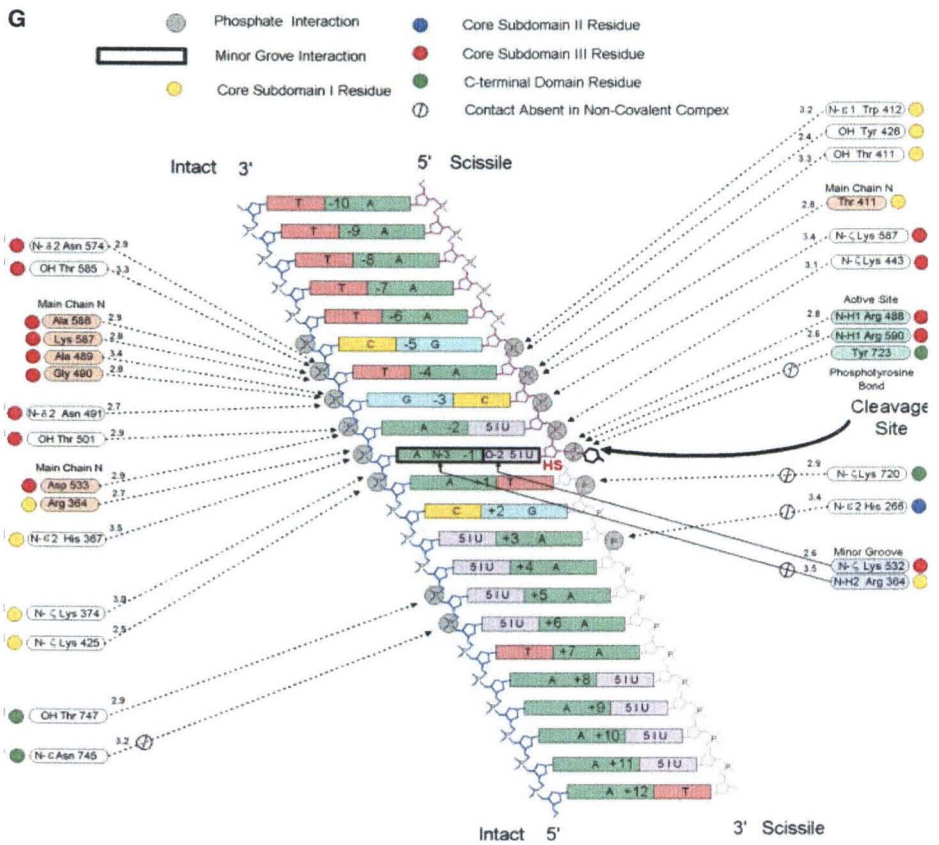

Core Subdomain I
Core Subdomain II
-10
+12
C-terminal Domain
Core Subdomain III
Cleavage Site

Figure 7.13. Schematic representation of the protein DNA interaction in the covalent complex of Topo I with DNA. Interactions between protein side chain and main chain atoms with DNA phosphate groups and minor groove atoms closer than 3.5 Å are indicated. Figure obtained from reference 6.


G
Phosphate Interaction
Minor Grove Interaction
Core Subdomain I Residue
Core Subdomain II Residue
Core Subdomain III Residue
C-terminal Domain Residue
Contact Absent in Non-Covalent Compex
Intact 3'
5' Scissile
Intact 5'
3' Scissile
Cleavage Site
Active Site
Main Chain N
Minor Groove
Phosphotyrosine Bond
HS

Preparation of Conjugates 1-3. ImPyPyPy-$(R)^{NH2}\gamma$-PyPyPyPy-β-Pam-Resin was synthesized in a stepwise fashion by solid phase methods according to established protocols.[19] Fmoc-6-aminocaproic acid (25.5 mg, 75 μmol) and HBTU (28 mg, 75 μmol) were added in 2 ml of DMF and 1 ml DIEA and shaken for 1 hour. The Fmoc group was deprotected with piperidine and the polyamide was cleaved off the resin with neat dimethylaminopropylamine in 16 hours. A solution of camptothecin acid **6** (3.1 mg, 7.5 μmol), N-hydroxysuccinimide (0.85 mg, 7.5 μmol), dicyclohexylcarbodiimide (16.7 mg, 81 μmol) in DMF (600 μl) was allowed to stir at room temperature for 2 hours. ImPyPyPy-γ^{CA5NH2}-PyPyPyPy-β-Dp (3.3 mg, 2.5 μmol) and diisopropylethylamine (50 μl) in DMF (650 μl) was added and allowed to stir at room temperature for 9 hours. The reaction was diluted with 0.1% trifluoroacetic acid in H_2O (6 ml) and purified by preparatory reverse phase HPLC. Lyophilization of the pure fractions afforded conjugate **1**. For the synthesis of conjugate **2,** Fmoc-6-aminocaproic acid was replaced with Fmoc-12-aminododecanoic acid. For the synthesis of conjugate **3**, Fmoc-6-aminocaproic acid was replaced with Boc 17-amino-4,7,10,13-tetraoxa-heptadecanoic acid. Yields and characterization: conjugate **1**: 0.4 mg, 10% recovery, MALDI-TOF [M+H] (monoisotopic), calcd 1753.90, obsd 1753; conjugate **2**: 0.5 mg, 13% recovery, MALDI-TOF [M+H] (monoisotopic), calcd 1838.06, obsd 1838.09; conjugate **3**: 1.6 mg, 24% recovery, MALDI-TOF [M+H] (monoisotopic), calcd 1916.08, obsd 1915.94.

ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-TA-Cpt **(7).** ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-Pam resin was synthesized in a stepwise fashion by Boc-chemistry manual solid phase protocols starting with a Boc-aminocaprylic acid monomer. A sample of resin was treated with neat 3,3'-diamino-N-methyldipropylamine (triamine-TA) (2 ml), heated (55°C, 24 hours) and purified by reversed phase HPLC. ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-TA was recovered as a white powder upon lyophilization of the appropriate fraction.

A solution of camptothecin acid **6** (3.1 mg, 7.5 μmol), N-hydroxysuccinimide (0.85 mg, 7.5 μmol), dicyclohexylcarbodiimide (16.7 mg, 81 μmol) in DMF (600 μl) was allowed to stir at room temperature for 2 hours. ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-TA (4.6 mg, 3.0 μmol) and diisopropylethylamine (50 μl) in DMF (650 μl) was added and allowed to stir at room temperature for 9 hours. The reaction was diluted with 0.1% trifluoroacetic acid in H_2O (6 ml) and purified by preparatory reverse phase HPLC. Lyophilization of the pure fractions afforded conjugate **7.** 2.5 mg, 43% recovery, MALDI-TOF-MS [M+H] (monoisotopic), calcd 1939.88 obsd 1940.0.

Preparation of Conjugates 8. ImImPyPy-$(R)^{NH_2}$γ-ImPyPyPy-β-Pam-Resin was synthesized in a stepwise fashion by solid phase methods according to established protocols.[19] Disuccinimidyl glutarate (100 mg, 306 μmoles) in 2 ml of DMF and 1 ml DIEA was added to the polyamide still on the resin and shaken for 1 hour followed by the addition of 4,9-dioxa-1,12-dodecanediamine (81 mg, 400 μmoles) and stirring for 5 hours. The polyamide was cleaved off the resin with neat dimethylaminopropylamine in 16 hours. A solution of camptothecin acid **6** (3.4 mg, 8.2 μmol), N-hydroxysuccinimide (0.95 mg, 8.2 μmol), dicyclohexylcarbodiimide (18.7 mg, 91 μmol) in DMF (400 μl) was allowed to stir at room temperature for 2 hours. ImImPyPy-γ^{NH_2}-ImPyPyPy-β-Dp (4.7 mg, 2.7 μmol) and diisopropylethylamine (130 μl) in DMF (200 μl) was added and allowed to stir at room temperature for 17 hours. The reaction was diluted with 0.1% trifluoroacetic acid in H_2O (6 ml) and purified by preparatory reverse phase HPLC. Lyophilization of the pure fractions afforded conjugate **8**. Yields and characterization: conjugate **8**: 0.5 mg, 9% recovery, MALDI-TOF [M+H] (monoisotopic), calcd 1942.88, obsd 1942.91.

Topoisomerase I cleavage experiment. Reaction mixtures (10 μl each) containing 50 mM Tris-HCl (pH 7.2), 50 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mM EDTA, 15 μg/ml BSA, 10 kcpm 3'-^{32}P-labeled 115 base pair pCW1 *Eco*RI/ *Afl*II

restriction fragment, were incubated with each polyamide-camptothecin conjugate **1-3** for 12 hours at 22°C. The Topoisomerase I reaction was initiated by the addition of 10 U of calf thymus Topo I followed by incubation at 37°C for 1 hour. The reaction was quenched by adding 10% SDS (1 μl) and proteinase K (1.5 mg/mL, 1.2 μl), followed by incubation at 50°C for 30 min. The reaction mixtures were ethanol precipitated and resuspended in 100 mM Tris-borate-EDTA/80% foramide loading buffer, denatured at 85°C for 10 min, placed on ice, and immediately loaded onto an 85°C denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 1.5 hours. The gels were dried under vacuum at 80°C and then quantitated using storage phosphor technology.

Acknowledgment. We are grateful to the National Institutes of Health (GM-27681) for research support and a traineeship award (GM-08501) to C.C.C.W.

Supporting Information Available. Quantitative DNase I footprint titration experiments with polyamide-camptothecin conjugate **2** and **3**. The material is available free of charge via the Internet at http://pubs.acs.org.

References

1. Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, 387, 202.
2. Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan; P. B. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 12890.
3. Dervan, P. B.; Burli, R. W. *Curr. Op. in Chem. Biol.* **1999**, 3, 688.
4. Dean, N. M.; McKay, R.; Condon, T. P.; Bennett, C. F. *J. Biol. Chem.* **1994**, 269, 16416.
5. Pommier, Y.; Pourquier, P.; Fan, Y.; Strumberg, D. *Biochim. Biophys. Acta-Gene Struct. Expression* **1998**, 1400, 83.
6. Redinbo, M. R.; Stewart, L.; Kuhn, P.; Champoux, J. J.; Hol, W. G. J. *Science* **1998**, 279, 1504.
7. Stewart, L.; Redinbo, M. R.; Qiu, X. Y.; Hol, W. G. J.; Champoux, J. J. *Science* **1998**, 279, 1534.
8. Hsiang, Y. H.; Hertzberg, R.; Hecht, S.; Liu, L. F. *J. Biol. Chem.* **1985**, 260, 4873.
9. Bendixen, C.; Thomsen, B.; Alsner, J.; Westergaard, O. *Biochemistry* **1990**, 29, 5613.
10. Matteucci, M.; Lin, K. Y.; Huang, T.; Wagner, R.; Sternbach, D. D.; Mehrotra, M.; Besterman, J. M. *J. Am. Chem. Soc.* **1997**, 119, 6939.
11. (a) Arimondo, P. B.; Bailly, C.; Boutorine, A.; Sun, J. S.; Garestier, T.; Helene, C. *Comptes Rendus Acad. Sci. Ser. III-Sci. Vie-Life Sci.* **1999**, *322*, 785-790. (b) Arimondo, P. B.; Moreau, P.; Boutorine, A.; Bailly, C.; Prudhomme, M.; Sun, J. S.; Garestier, T.; Helene, C. *Bioorg. Med. Chem.* **2000**, *8*, 777-784.
12. Zhao, R. L.; AlSaid, N. H.; Sternbach, D. L.; Lown, J. W. *J. Med. Chem.* **1997**, *40*, 216-225.

13. Fan, Y.; Weinstein, J. N.; Kohn, K. W.; Shi, L. M.; Pommier, Y. *J. Med. Chem.* **1998**, 41, 2216.

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

15. Kingsbury, W. D.; Boehm, J. C.; Jakas, D. R.; Holden, K. G.; Hecht, S. M.; Gallaher, G.; Caranfa; M. J.; McCabe, F. L.; Faucette, L. F.; Johnson, R. K.; Hertzberg, R. P. *J. Med. Chem.* **1991**, 34, 98.

16. a) Wani, M. C.; Ronman, P. E.; Lindley, J. T.; Wall, M. E. *J. Med. Chem.* **1980**, 23, 554. b) Wall M. E.; Wani, M. C.; Nicholas, A. W.; Manikumar, G.; Tele, C.; Moore, L.; Truesdale, A.; Leitner, P.; Besterman, J. M. *J. Med. Chem.* **1993**, 36, 2689.

17. Jaxel, C.; Kohn, K. W.; Pommier, Y. *Nucleic Acids Res.* **1988**, 16, 11157.

18. Edwards, K. A.; Halligan, B. D.; Davis, J. L.; Nivera, N. L.; Liu, L. F. *Nucleic Acids Res.* **1982**, 10, 2565.

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

20. Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, 120, 1382.

21. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, 97, 3930.

Chapter 8

Application of Camptothecin Polyamide Conjugates to Study DNA Repair and Recombination

Abstract

Pyrrole-Imidazole (Py-Im) polyamides are cell permeable ligands that can be targeted to any predetermined DNA sequence. Polyamide camptothecin conjugates are capable of cleaving DNA in high yields at T^G sites proximal to the polyamide binding site by trapping the endogenous enzyme Topo I. Because of the generality of the pairing rules, and polyamides demonstrated utility in *in vivo* systems, these conjugates may be useful for studying the effects of this class of DNA binding ligands on homologous recombination and DNA repair. Described are the design and synthesis of camptothecin polyamide conjugates that cleave DNA at a determined site on a specific gene. Currently, these compounds are being studied *in vivo* in hopes that DNA cleavage will stimulate a DNA repair event and homologous recombination. If successful, these studies could have future applications in the areas of gene therapy.

Introduction

Pyrrole imidazole polyamides that can target specific sequences of DNA and bind with high affinity have been shown *in vitro* and *in vivo* to inhibit or activate transcription by competing with the binding of required transcription factors.[1] By attaching activator peptides, Im/Py polyamides are capable of recruiting the transcription machinery *in vitro*.[2] The question arises whether polyamides could be used to study gene repair.

To study DNA repair and recombination, we proposed using Py-Im polyamides to create sequence specific damages to the DNA. The use of polyamide conjugates allows us to control the location of the damage inflicted in order to stimulate repair and recombination to the particular gene in study. In collaboration with Dr. Matt Porteus of the Baltimore group at Caltech, experiments were designed to use camptothecin polyamide conjugates to trap Topoisomerase I and create sequence specific damages to the DNA.[3] In addition to using camptothecin, *seco*-CBI (Dr. Aileen Chang of the Dervan group) and nitrogen mustards (Mr. Nicholas Wurtz of the Dervan group) are also currently being studied. Described here are the efforts in using camptothecin polyamide conjugates for these experiments.

Experimental Design

The experimental design was based on observations in the Baltimore group using Green Fluorescent Protein (GFP). When a mutant GFP gene with an internal stop codon is expressed, functional GFP is not detected (Figure 8.1a). When substrate DNA is added containing the wild type gene, homologous recombination occurs such that functional GFP can be detected for 1 in every 10^6 cells (Figure 8.1b). If instead of the internal stop codon, a site for a rare-cutting restriction enzyme is inserted[4], a double strand break can be generated. For the case of homologous recombination with the substrate for the double strand break, functional GFP can be detected for 1 in every 500 cells (Figure 8.1c). Thus these experiments were designed to address the question of

whether or not polyamides could be used to incorporate a DNA strand break, and if so could we generate similar results of recombination and mutagenesis (Figure 8.1d).

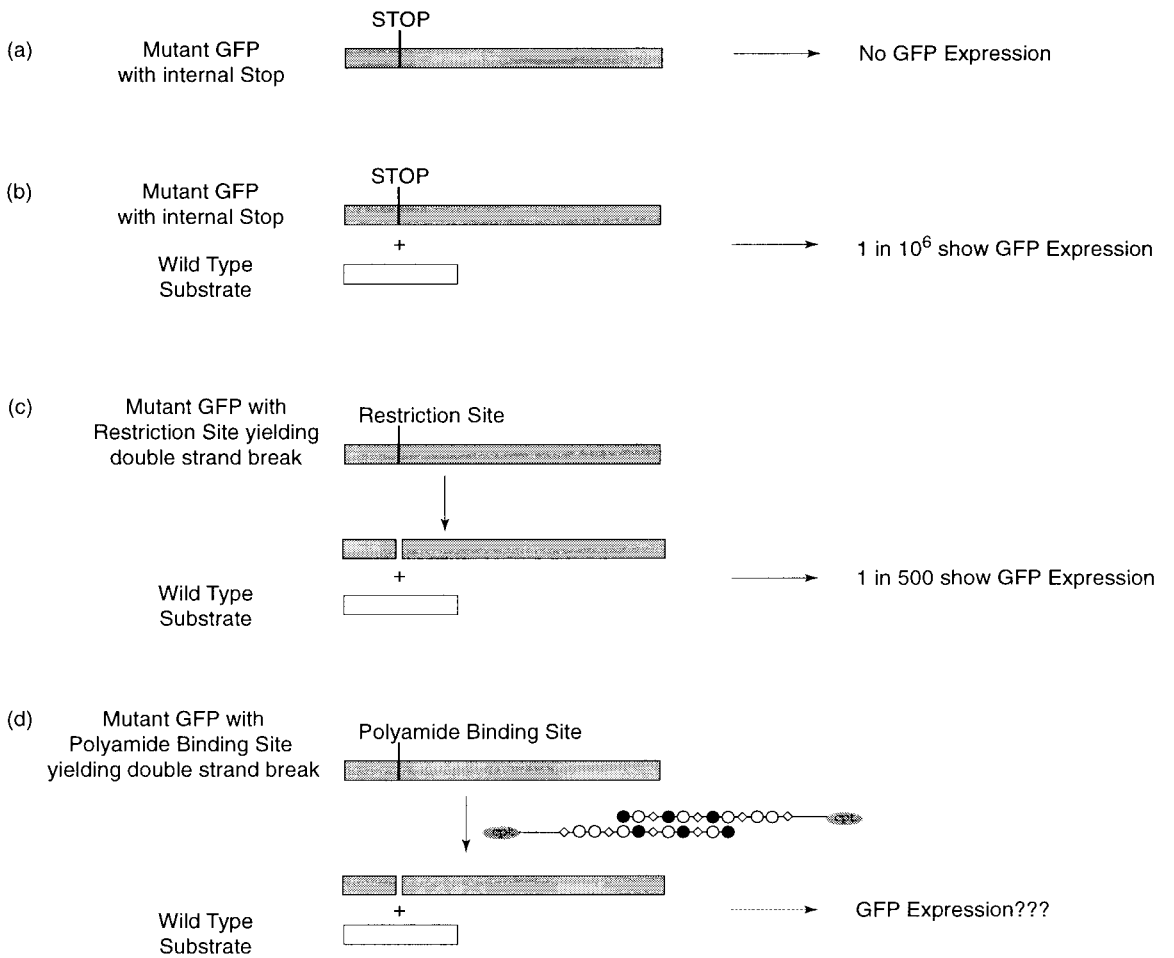
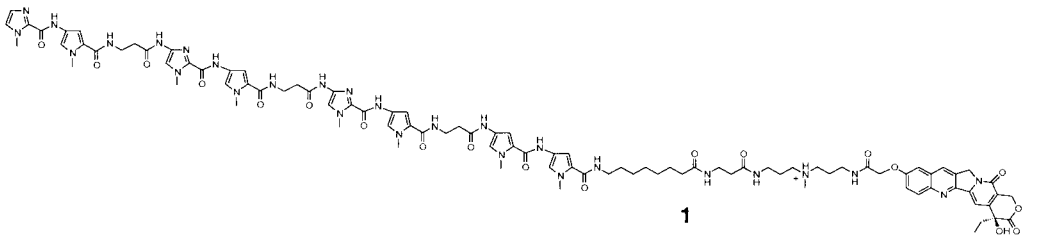


Figure 8.1. Experimental design for proposed polyamide conjugates. See text for further details.

High affinity, high specificity, and large binding site size are criteria for the selection of the polyamide portion of the conjugate. A large binding site would decrease the number of match site for the conjugate in the genome. The eight ring polyamide ImPy-β-ImPy-β-ImPy-β-PyPy-β-Dp capable of binding 16 contiguous base pairs 5'-

ATAAGCAGCTGCTTTT-3' as a homodimer was chosen for the polyamide portion of conjugate **1**.[5] ImPy-β-ImPy-β-ImPy-β-PyPy-β-Dp binds its match site with an affinity of 3.5 x 10^{10} M^{-1} and a ten-fold specificity toward mismatches. A tandem hairpin conjugate ImPyPy-(*R*)[ImPyPy-(*R*)H_2Nγ-PyPyPyδ]^{HN}γ-PyPyPy-β-Dp binding the 11 base pair sequence 5'-TGTTATTGTTA-3' with an affinity of greater than 1.0 x 10^{12} M^{-1} was chosen for the polyamide portion of conjugate **2**.[6] ImPyPy-(*R*)[ImPyPy-(*R*)H_2Nγ-PyPyPyδ]^{HN}γ-PyPyPy-β-Dp binds its match site with greater than 4500-fold specificity over double mismatches. Both parent compounds were chosen because they match the criteria chosen for this study. Tandem conjugate **2** by itself will cause a single strand break but it is known that the single strand break will be transformed into a double strand break by both the replication and transcription machinery. The insert containing the binding site of the polyamide for each construct (homodimer and tandem) was inserted in the middle of the GFP gene (Figure 8.3). The sequence of the binding sites in the insert is identical to published result with the exception that known camptothecin polyamide sites are inserted next to the polyamide site. The insert includes a stop codon so that cells that do not undergo recombination will not express functional GFP. In the event of a double strand break, repair processes could correct the mutation, which would then be detectable by fluorescence.



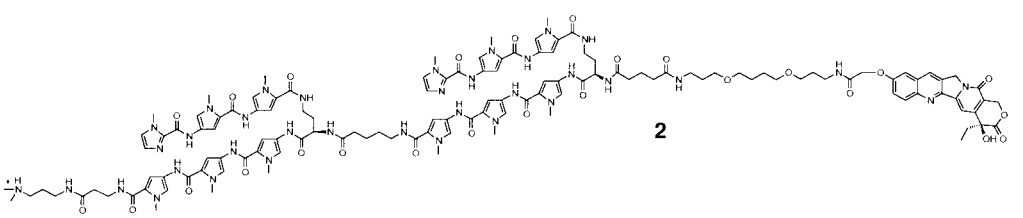


Figure 8.2. Structures of the polyamide conjugates used for this study.

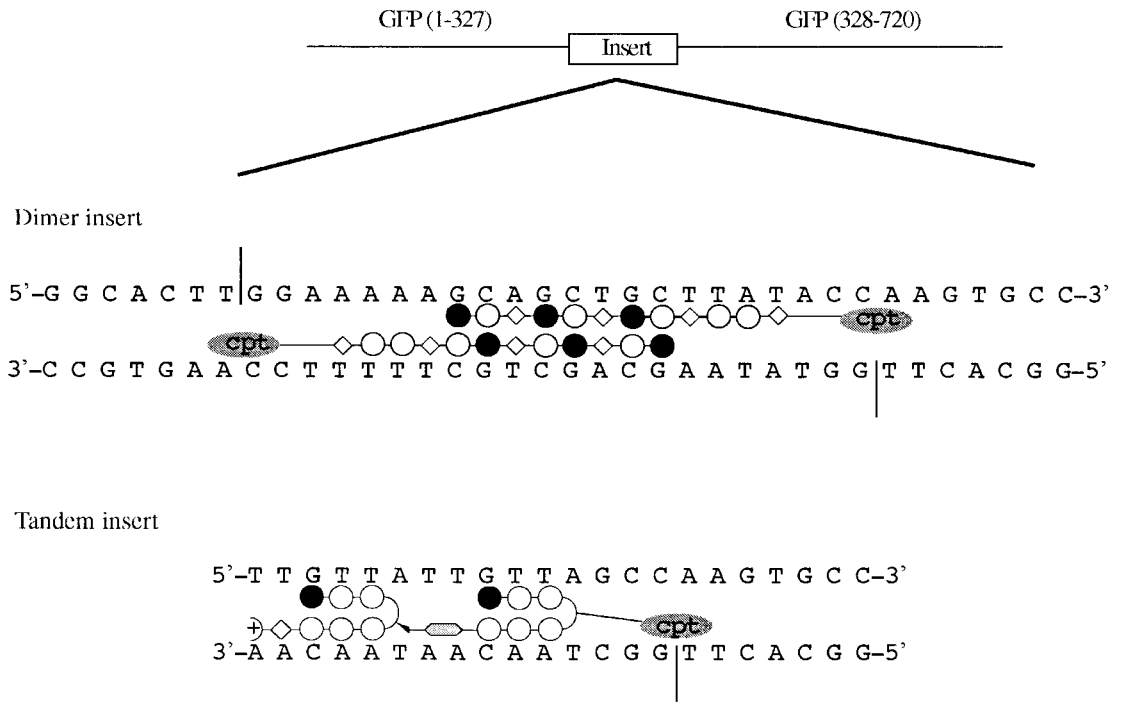


Figure 8.3. DNA constructs of homodimer and tandem inserts for GFP experiments.

Results

Synthesis of camptothecin polyamide conjugate

The parent unmodified polyamides **1** and **2** were synthesized on solid support as previously reported. Polyamide **1** was cleaved from the resin with 3,3'-diamino-N-methyldipropylamine (triamine, TA) and the conjugates were then synthesized by addition of the camptothecin acid activated with HOBt and DCC. Polyamide **2** was synthesized with (*R*)-2,4-diaminobutyric acid (DABA) in the turn position using Boc chemistry machines assisted protocol. A polyethylene linker was added to the amine of the turn and subsequent addition of activation of camptothecin acid yielded the desired camptothecin polyamide conjugate. Synthesis of conjugate **1** is shown in Figure 8.4.

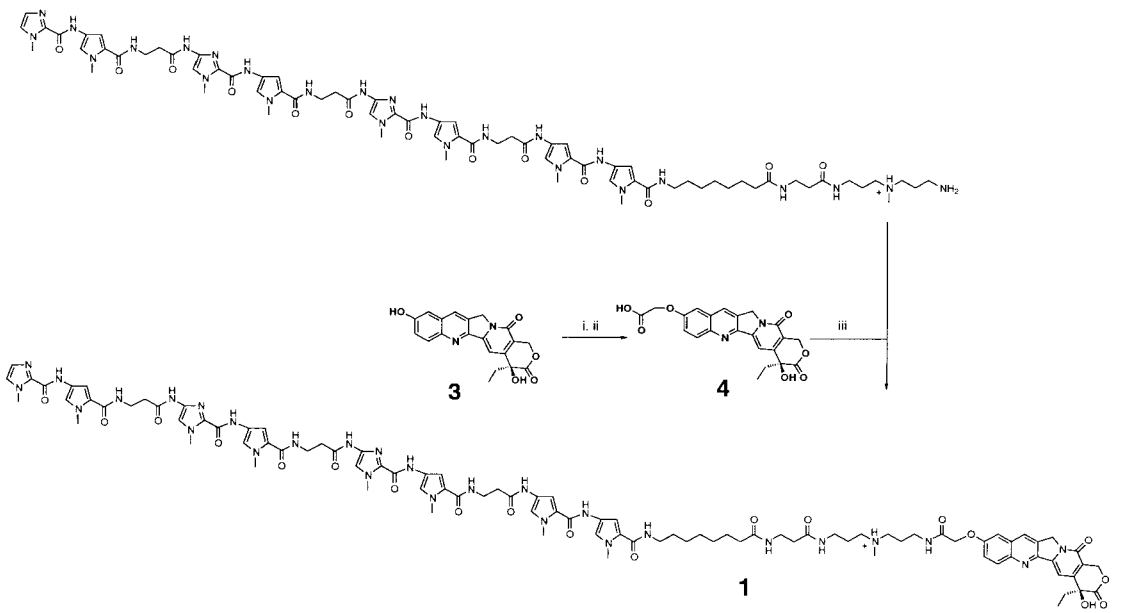


Figure 8.4. Synthesis of polyamide camptothecin conjugate **1**.

Topoisomerase I cleavage assay

Polyamide **1** was analyzed on a 221 base pair fragment of GFP containing the designed insert (Figure 8.5a). The DNA fragments are 5' ^{32}P labeled by PCR for technical reasons. PCR labeling allows any section of the DNA fragment to be amplified and labeled without the need for restriction enzyme sites unlike labeling with 3' fill in. However, Topo I cleavage causes the Topo I enzyme to be physically associated with the 5' labeled fragment. To degrade the Topo I enzyme, the fragments are treated with Proteinase K. Even with the protease treatment the exact location of the cleavage could only be estimated by comparison to the sequencing reaction. For both conjugates, the cleavage site is the same as the camptothecin control. Since the camptothecin cleavage site was artificially placed next to the polyamide binding site, the cleavage site should be the T^G site 3 to 4 base pairs removed from the designed binding site. At 10 nM concentrations, **1** cleaves this match site with a yield of 30% on the top strand (Figure 8.6). On the bottom strand cleavage was seen at 60%. Polyamide **2** was assayed on a 215 base pair fragment containing the tandem polyamide match site (Figure 8.5b). The tandem conjugate cleaves this match site with a yield of 50% (Figure 8.7).

a) GFP Dimer Fragment (177 bp+ 44 bp-insert)

```
5'-AAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC
3'-TTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTCGCGTGGTAGAAGAAG

AAGGACGACGGCAACTACAAGACC-TAAGCTCTCGTGGCACTTGGTATAAGCAGCTGCTTTTTCCAAGT
TTCCTGCTGCCGTTGATGTTCTGG-ATTCGAGAGCACCGTGAACCATATTCGTCGACGAAAAAGGTTCA

GCCCTCGAGCTT-CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGG
CGGGAGCTCGAA-GCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCC

CATCGACTTCAAGGAGGACGGCAACATCCTG-3'
GTAGCTGAAGTTCCTCCTGCCGTTGTAGGAC-5'
```

b) GFP Tandem Fragment (177 bp+ 38 bp-insert)

```
5'-AAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC
3'-TTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCCTCGCGTGGTAGAAGAAG

AAGGACGACGGCAACTACAAGACC-TAAGCTTGGCACTTGGGTAACAATAACAACTCGAGCTT-CGCGC
TTCCTGCTGCCGTTGATGTTCTGG-ATTCGAACCGTGAACCCATTGTTATTGTTGAGCTCGAA-GCGCG

CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA
GCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGCTGAAGTTCCTCCT

CGGCAACATCCTG-3'
GCCGTTGTAGGAC-5'
```

Figure 8.5. GFP fragments used in this study. Insert with polyamide binding sites are between the dashes, polyamide binding sites are in bold. (a) dimer construct, (b) tandem construct.

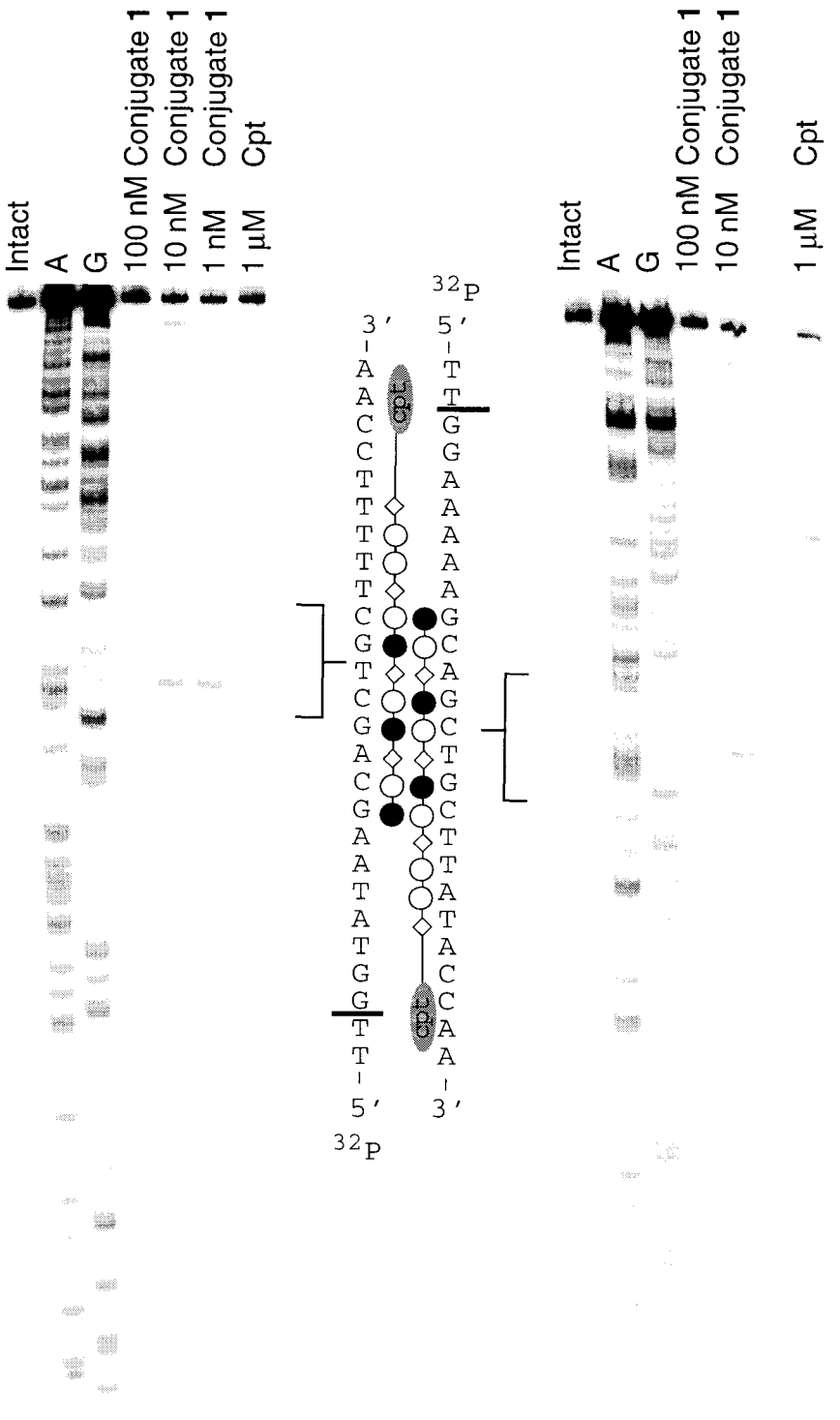


Figure 8.6. Topo I cleavage experiment with the dimer conjugate **1.**

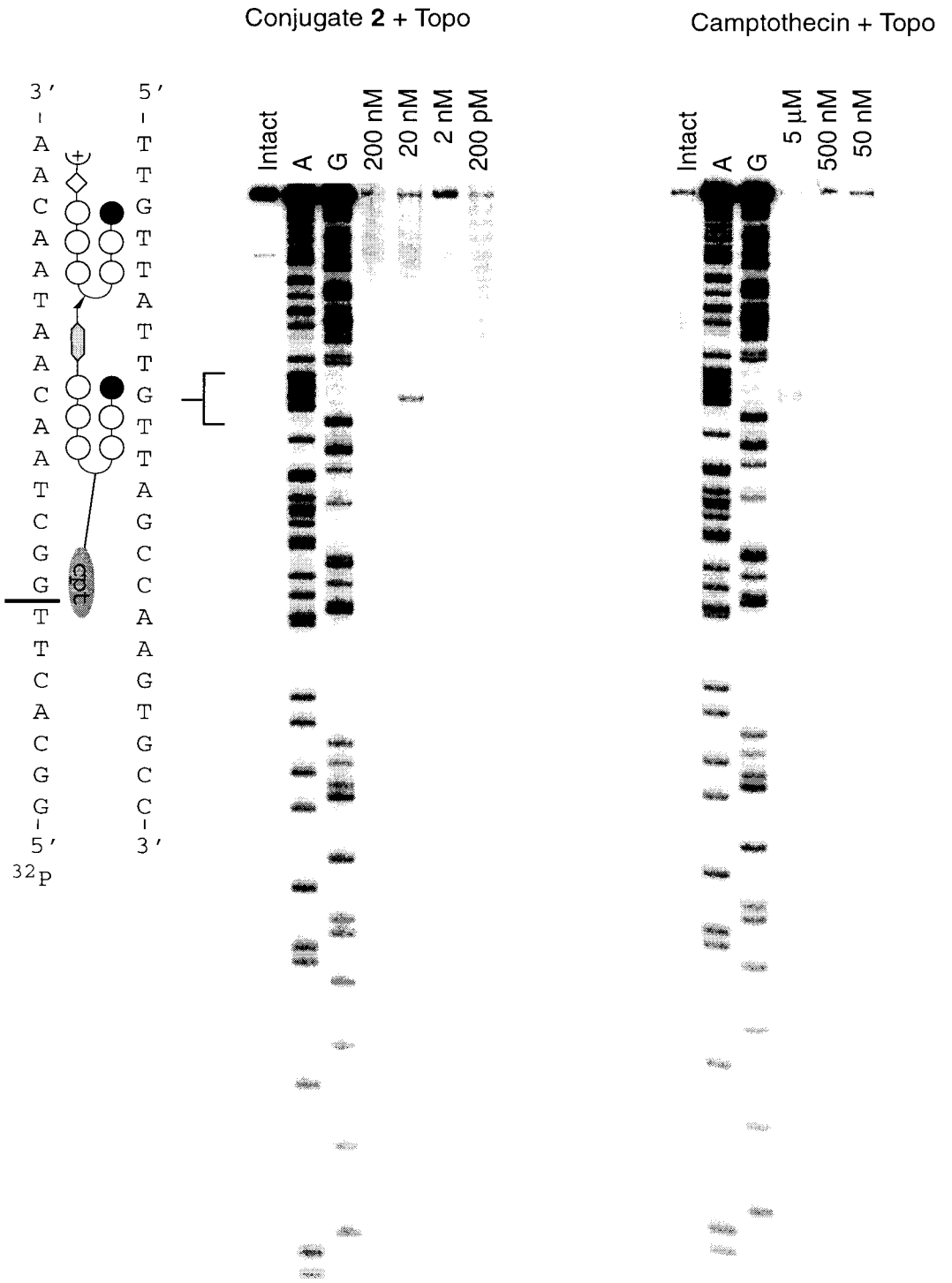


Figure 8.7. Topo I cleavage experiment with tandem conjugate **2**.

Gene Correction Experiments

Currently, Dr. Matt Porteus of the Baltimore group is working to set up these experiments to study the effects of Py-Im polyamides on homologous recombination. It is hoped that upon treatment with polyamide-camptothecin conjugates, cells that had undergone recombination or repair would be detected by expression of GFP. Preliminary results have shown that the polyamide-camptothecin conjugates are toxic to cells at 1 μM but not 100 nM while the camptothecin control is toxic to cells at 100 nM but not 10 nM. The difference in the toxicity of conjugate versus camptothecin could be due to decrease in cell permeability of the conjugate. Cell permeation of polyamides is currently not very well understood and studies are underway in the Dervan group including using nuclear localization sequences to traffic polyamides into the nucleus. Alternatively the difference in toxicity could be due to the sequence specificity of the conjugate. Maybe by being more selective in cleaving the DNA the conjugate is less toxic. Besides looking at toxicity difference, gene specific reduction in RNA synthesis was also tested. Preliminary experiments showed reduction in global levels RNA production but no gene specific reduction in RNA synthesis was observed.

Conclusions

The design of Py-Im polyamide conjugates with camptothecin incorporates the ability to target specific DNA sequences and traps Topo I onto the DNA in a sequence specific fashion. We believe that this class of molecules will provide powerful tools for studying and manipulating transcription in the cell. The results of the forthcoming *in vivo* experiments will shape the design of additional motifs for these classes of compounds and point to new directions to apply these molecules.

Experimental

Materials

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. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 × 100 mm, 300 Å C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/minutes.

Synthesis camptothecin polyamide conjugates

ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-TA-Cpt **(1).** ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-Pam resin was synthesized in a stepwise fashion by Boc-chemistry manual solid phase protocols starting with a Boc-aminocaprylic acid monomer. A sample of resin was treated with neat 3,3'-diamino-N-methyldipropylamine (triamine-TA) (2 ml), heated (55°C, 24 hours) and purified by reversed phase HPLC. ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-TA was recovered as a white powder upon lyophilization of the appropriate fraction. A solution of camptothecin acid **4** (3.1 mg, 7.5 μmol), N-hydroxysuccinimide (0.85 mg, 7.5 μmol), dicyclohexylcarbodiimide (16.7 mg, 81 μmol) in DMF (600 μl) was allowed to stir at room temperature for 2 hours. ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-TA (4.6 mg, 3.0 μmol) and diisopropylethylamine (50 μl) in DMF (650 μl) was added and allowed to stir at room temperature for 9 hours. The reaction was diluted with 0.1% trifluoroacetic acid in H_2O (6 ml) and purified by preparatory reverse phase HPLC. Lyophilization of the pure fractions afforded conjugate **1.** 2.5 mg, 43% recovery, MALDI-TOF-MS [M+H] (monoisotopic), calcd 1939.88 obsd 1940.0.

ImPyPy-(*R*)[ImPyPy-(*R*)^{Cpt}γ-PyPyPyδ]^{HN}γ-PyPyPy-β-Dp **(2).** ImPyPy-(*R*)[ImPyPy(*R*)^{H_2N}γ-PyPyPyδ]^{HN}γ-PyPyPy-β-Pam resin was synthesized in a stepwise fashion by Boc-chemistry machine assisted solid phase. A sample of resin was treated

with neat (dimethylamino)-propylamine (2 ml), heated (55°C, 24 hours) and purified by reversed phase HPLC. ImPyPy-(*R*)[ImPyPy-(*R*)^{Cpt}γ-PyPyPyδ]^{HN}γ-PyPyPy-β-Dp was recovered as a white powder upon lyophilization of the appropriate fraction. A solution of camptothecin acid **4** (3.1 mg, 7.5 μmol), N-hydroxysuccinimide (0.85 mg, 7.5 μmol), dicyclohexylcarbodiimide (16.7 mg, 81 μmol) in DMF (600 μl) was allowed to stir at room temperature for 2 hours. ImPy-β-ImPy-β-ImPy-β-PyPy-β-C7-TA (3.3 mg, 2.5 μmol) and diisopropylethylamine (50 μl) in DMF (650 μl) was added and allowed to stir at room temperature for 9 hours. The reaction was diluted with 0.1% trifluoroacetic acid in H_2O (6 ml) and purified by preparatory reverse phase HPLC. Lyophilization of the pure fractions afforded conjugate **2.** (14.4 mg, 10.4 μmoles, 6.8% recovery). UV (H_2O) λ_{max} (ε), 312 nm, (104,250); MALDI-TOF-MS (monoisotopic) [M+H] 1381.70 (calculated 1380.64 for $C_{62}H_{80}N_{26}O_{12}$).

DNA Reagents and Materials. Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine 5'-[γ-^{32}P] triphosphates were purchased from I.C.N. Sonicated, deproteinized calf thymus DNA was acquired from Pharmacia. RNase free water was obtained from USB and used for all reactions. All other reagents and materials were used as received. All DNA manipulations were performed according to standard protocols.

Construction of Plasmid DNA. The plasmids GFP1 and GFP2 were constructed by Dr. Matt Porteus. Concentration of the prepared plasmid was determined at 260 nm from the relationship of 1 OD unit=50 μg mL^{-1} duplex DNA.

PCR Labeling to generate 5'-End-Labeled Restriction Fragments. Two 21 mer primers were synthesized for PCR amplification: primer A (top)5'-AAG-CAG-CAC-GAC-TTC-TTC-AAG-3' and primer B (bottom) 5'-CAG-GAT-GTT-GCC-GTC-CTC-CTT-3'. To label the top strand, primer A was treated with T4 polynucleotide kinase and deoxyadenosine 5'-[γ-^{32}P] triphosphate as previously described. PCR reactions containing 60 pmol each primer, 10 μl PCR buffer (Boehringer-Mannheim), 3.7 μl

template (0.003 μg/mL), 2 μl dNTP mix (each at 10 mM), 1 μl 100X BSA (New England Biolabs) and 83 μl water were heated at 70°C for 5 minutes. Four units of Taq Polymerase were added (Boehringer-Mannheim). Thirty amplification cycles were performed, each cycle consisting of the following segments: 94°C for 1 minute, 54°C for 1 minute, and 72°C for 1.5 minutes. Following the last cycle, 10 minutes of extension at 72°C completed the reaction. The PCR products were gel purified as previously reported for 3'-end labeling protocols. To label the bottom strand, primer B was labeled.

Topoisomerase I cleavage experiment. Reaction mixtures (10 μl each) containing 50 mM Tris-HCl (pH 7.2), 50 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mM EDTA, 15 μg/ml BSA, 10 kcpm 5'-^{32}P-labeled 115 bp pCW1 *Eco*RI/ *Afl*II restriction fragment, were incubated with each polyamide-camptothecin conjugate **1** and **2** for 4 hours. The Topoisomerase I reaction was initiated by the addition of 10 U of calf thymus Topo I followed by incubation at 37°C for 1 hour. The reaction was quenched by adding 10% SDS (1 μl) and proteinase K (1.5 mg/mL, 1.2 μl), followed by incubation at 50°C for 30 minutes. The reaction mixtures were ethanol precipitated and resuspended in 100 mM Tris-borate-EDTA/80% foramide loading buffer, denatured at 85°C for 10 minutes, placed on ice, and immediately loaded onto an 85°C denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 1.5 hours. The gels were dried under vacuum at 80°C and then quantitated using storage phosphor technology.

Acknowledgements

We are grateful to the National Institutes of Health for research support and the Ralph M. Parsons foundations for a predoctoral fellowship to C.C.C.W. We thank Dr. Matt Porteus of the Baltimore group at Caltech and Dr. Aileen Chang and Mr. Nicholas Wurtz of the Dervan group for their collaborative efforts on this project. We thank G.M. Hathaway

and the Caltech Protein/Peptide Microanalytical Laboratory for MALDI-TOF mass spectrometry.

References

1. Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan; P. B. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 12890.
2. Mapp, A. K.; Ansari, A. Z.; Ptashne, M; Dervan, P. B. Activation of gene expression by small molecule transcription factors. *Proc. Natl. Acad. Sci. USA* **2000,** 97, 3930-3935.
3. Wang C. C. W.; Dervan, P. B. *J. Am. Chem. Soc.* **2001,** in press.
4. Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* **12**, 224-228 (1996).
5. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, 120, 3534-3535.
6. Herman D. M.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **2000**, 24, 4487-4497.