

**SEQUENCE-SPECIFIC MINOR GROOVE BINDING POLYAMIDES:
DNA RECOGNITION AND APPLICATIONS**

Thesis by
Scott Reynolds Carter
In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1999

(Submitted December 16, 1998)

© 1999

Scott Reynolds Carter

All Rights Reserved

Acknowledgements

I would like to thank Prof. Peter Dervan for his guidance through these projects.

In my time at Caltech, I have been fortunate to work with a number of excellent post-docs in the Beckman Institute. I am especially indebted to Yitzhak Tor, for his patient teaching in my first year. Darren Magda, Ken Turnbull, Andrea Staubli, Steve Woski, Carol Wada, Jon Parquette, Thomas Lehmann, Robert Hudson, Kathy Grant, Paul Floreancig, and Anna Mapp all provided assistance and friendship over the years.

I thank Margot Hoyt, Steve Gould, Chris Smith, and Dian Buchness for their kind assistance over the years. It is truly appreciated.

Michelle Arkin and Shana Kelley shared the joys and burdens of being graduate students in the Beckman Institute with me. I thank them especially for their long years of HPLC maintenance.

I am most grateful to the current Beckman Institute contingent for their great camaraderie and tolerance during the completion of this work, especially to Simon Friedman and Ralf Jäger for late night help. In particular, I would also like to thank Roland Bürli for the exciting collaboration we undertook toward the study of the recognition of G-rich sequences, and for his thorough editing of this manuscript.

I thank all of my other friends who have made life at Caltech more enjoyable. Helmut Brunar and Jean-Francois Mouscadet taught me never again to trust a European claiming to be a “pretty good” skier. I thank Ken Brameld for forcing me to take advantage of the rare and wonderful blizzards at Mt. Baldy. Andrew and Stacy Merickel provided many inspirational meals.

I would especially like to acknowledge my family for their support during this endeavor.

Finally, I thank Stephanie Teleki for her love and friendship through this process.

Abstract:

Pyrrole-imidazole polyamides bind to the minor groove of DNA with high affinity and specificity. This thesis explores the scope of this DNA recognition motif both in terms of improved recognition and application to problems of DNA detection and purification. Chapter 2 describes conjugates of polyamides with fluorescent dyes. Three cyanine dye butyric acids were synthesized and linked to the polyamide, ImPyPyPy- γ -ImPyPyPy- β -Dp. These conjugates retain high specificities and high affinities for targeted DNA sequences, and show large fluorescence enhancements upon binding to double stranded DNA. Chapter 3 describes the use of polyamide conjugates as affinity reagents for the purification of double stranded DNA. Specifically, systems composed of biotin and histidine tagged conjugates that can rapidly and efficiently capture restriction fragment DNA are described.

Chapter 4 focuses on the design and synthesis of constrained γ -aminobutyric acid analogs designed to improve hairpin polyamide binding affinities through reduction of the entropic cost of binding to DNA. Six derivatives were synthesized as Boc protected monomers and were incorporated into polyamides by solid phase synthesis. The affinity and specificity of the compounds were analyzed by quantitative DNase I footprinting.

Chapter 5 describes preliminary synthetic efforts towards the recognition of G rich DNA sequences with polyamides containing imidazole dimers linked by alkyl chains rather than amides. Regioselective synthesis of an alkyl imidazole dimer suitably protected for solid phase polyamide synthesis is described. This dimer was incorporated into 2 polyamides for the recognition of 5'-WGGGGW-3' sequences. Alkyl substituted polyamides bound comparably to the corresponding amide linked compounds.

An appendix summarizes work on the recognition of 5'-WGWGW-3' sequences by pyrrole-imidazole polyamides. ImPyIm- γ -PyPyPy- β -Dp and ImPyImPy- γ -PyPyPyPy- β -Dp recognize a 5'-TGTGTA-3' binding site.

Chapter 6 describes studies of the non-natural base pair between isoguanosine and isocytosine. A synthesis of deoxyisoguanosine (d-isoG) suitably protected for solid phase oligonucleotide synthesis is presented. Transcriptional studies on templates containing d-isoG showed that it directed the transcription of uridine rather than isoC, providing evidence that isoG can access different tautomeric forms.

Table of Contents

Chapter 1. Introduction to DNA Recognition

1.1 BIOLOGICAL ROLES OF NUCLEIC ACID RECOGNITION.....	1
1.1.1 DNA STRUCTURE.....	1
1.1.2 TRANSCRIPTION REGULATION	4
1.2 PROTEIN AND MACROMOLECULE DNA INTERACTIONS.....	6
1.2.1 DNA BINDING MOTIFS.....	6
1.2.2 DESIGN OF SEQUENCE-SPECIFIC DNA BINDING PROTEINS	8
1.2.3 DNA RECOGNITION BY TRIPLE HELIX FORMATION	9
1.2.4 DNA RECOGNITION BY POLYAMIDE NUCLEIC ACID (PNA).....	11
1.3 SMALL MOLECULE-DNA INTERACTIONS.....	12
1.3.1 INTERCALATION.....	12
1.3.2 GROOVE BINDING	14
1.3.3 SIMULTANEOUS GROOVE BINDING AND INTERCALATION.....	16
1.3.3 DISTAMYCIN ANALOGS	16
1.3.4 HYBRID BINDING MOECULES.....	17
1.3.5 FUNCTIONAL DISTAMYCIN CONJUGATES	18
1.4 MINOR GROOVE BINDING POLYAMIDES.....	18
1.4.1 DIMERIC BINDING.....	19
1.4.2 PAIRING RULES FOR POLYAMIDE BINDING.....	20
1.4.3 POLYAMIDE MOTIFS	22
1.4.4 SYNTHESIS OF POLYAMIDES	23
1.4.5 POLYAMIDE CONJUGATES	24
1.5 THIS WORK.....	25

1.6 REFERENCES.....	26
----------------------------	-----------

Chapter 2: Sequence-Specific Detection of Double-Stranded DNA

2.1 ABSTRACT.....	32
--------------------------	-----------

2.2 DETECTION WITH CYANINE DYE CONJUGATES.....	32
---	-----------

2.2.1 BACKGROUND	32
------------------------	----

2.2.2 DESIGN	35
--------------------	----

2.2.3 SYNTHESIS OF COMPOUNDS.....	36
-----------------------------------	----

2.2.4 RESULTS AND DISCUSSION	42
------------------------------------	----

2.2.5 CYANINE CONJUGATE CONCLUSIONS	55
---	----

2.3 TOWARDS DETECTION THROUGH ENERGY TRANSFER.....	56
---	-----------

2.3.1 BACKGROUND	56
------------------------	----

2.3.2 DESIGN	58
--------------------	----

2.3.3 SYNTHESIS OF COMPOUNDS.....	59
-----------------------------------	----

2.3.4 RESULTS AND DISCUSSION	61
------------------------------------	----

2.3.5 DNA BINDING OF POLYAMIDE-FLUOROPHORE CONJUGATES.....	66
--	----

2.3.6 ENERGY TRANSFER CONCLUSIONS	67
---	----

2.4 DETECTION THROUGH PYRENE EXCIMER FORMATION.....	69
--	-----------

2.4.1 BACKGROUND	69
------------------------	----

2.4.2 DESIGN OF PYRENE CONJUGATES.....	69
--	----

2.4.3 RESULTS AND DISCUSSION	72
------------------------------------	----

2.5 CONCLUSIONS AND FUTURE OUTLOOK.....	75
--	-----------

2.5.1 GENOMIC ANALYSIS	75
------------------------------	----

2.5.2 BIOPHYSICAL ANALYSIS	76
2.5.3 CELLULAR STUDIES.....	77
2.6 EXPERIMENTAL.....	78
2.7 REFERENCES.....	84

Chapter 3: Polyamide Conjugates for Affinity Purification of DNA

3.1 BACKGROUND	87
3.2 SYNTHESIS OF POLYAMIDE CONJUGATES	89
3.3 RESULTS AND DISCUSSION.....	94
3.4 CONCLUSIONS.....	109
3.5 EXPERIMENTAL.....	109
3.5.1 MATERIALS.....	109
3.5.2 GENERAL PROCEDURES	110
3.6 REFERENCES.....	114

Chapter 4: γ -Aminobutyric Acid Analogs for Polyamide Hairpin Formation

4.1 BACKGROUND	115
4.2 SYNTHESIS OF PROTECTED TURN AMINO ACIDS.....	118

4.3 RESULTS AND DISCUSSION.....	122
4.4 CONCLUSIONS.....	128
4.5 EXPERIMENTAL.....	130
4.7 REFERENCES.....	139

Chapter 5: Recognition of G-Rich DNA Sequences

5.1 BACKGROUND	141
5.2 SYNTHESIS	144
5.3 DNA BINDING	152
5.4 CONCLUSIONS.....	153
5.5 EXPERIMENTAL.....	155
5.6 REFERENCES.....	163

Chapter 6: Studies on the *isoG-isoC* Base Pair

6.1 ABSTRACT.....	165
6.2 BACKGROUND	165
6.3 APPLICATION OF THE D- ^{ME} ISOC-AH-ISOG SYSTEM TO THE STUDY OF THE HAMMERHEAD RIBOZYME	170

6.4 STUDIES ON THE TRANSCRIPTION OF TEMPLATES CONTAINING DEOXY- ISOGUANOSINE.....	177
6.5 CONCLUSIONS.....	182
6.6 EXPERIMENTAL SECTION.....	185
6.7 REFERENCES.....	195

Appendix: Related Polyamide Studies

A.1 BACKGROUND.....	199
A.2 RECOGNITION OF 5'-WGWW-3' SEQUENCES.....	199
A.3 SYNTHESIS OF AN FMOC-AMINOPROPYL DERIVATIVE OF PYRROLE....	206
A.4 SYNTHESIS OF OTHER PYRROLE DERIVATIVES	208
A.5 EXPERIMENTAL.....	209
A.6 REFERENCES.....	212

List of Illustrations

Chapter 1. Introduction to DNA Recognition

Figure 1. Chemical Structure of DNA.....	2
Figure 2. Hydrogen bonding in DNA base pairs	3
Figure 3. Model of B-form DNA	4
Figure 4. Central dogma of molecular biology.....	4
Figure 5. Model of the HIV-1 promoter.	6
Figure 6. Common transcription factor DNA binding motifs.....	7
Figure 7. Minor groove binding transcription factors	8
Figure 8. Base triplet structures for pyrimidine and purine motif triple helices.....	10
Figure 9. PNA structure.....	12
Figure 10. Intercalative DNA binding compounds.....	13
Figure 11. DNA binding bis-intercalators:.....	14
Figure 12. Minor groove binding compounds.....	15
Figure 13. Ligands that interact through intercalation and groove binding.....	16
Figure 14. Structure of 1:1 complex of distamycin with DNA.....	17
Figure 15. Schematic polyamide representation.....	19
Figure 16. Pairing rules for DNA recognition by polyamides.....	21
Figure 17. X-ray crystal structure of ImHpPyPy- β -Dp	22

Chapter 2: Sequence-Specific Detection of Double-Stranded DNA

Figure 1. Conformation and resonance of thiazole orange.....	34
Figure 2. Model of ImPyPy- γ -PyPyPy- β -Dp-TO.....	36
Figure 3. Synthesis of cyanine dye derivatives.....	38
Figure 4. Synthesis of cyanine dye conjugates.....	41
Figure 5. Structures of the polyamide cyanine dye conjugates used in this study.	42

Figure 6. Absorbance spectra of thiazole orange.....	43
Figure 7. Absorbance spectra of thiazole yellow and thiazole blue derivatives.....	44
Figure 8. Oligonucleotides and duplexes used for fluorescence experiments.....	45
Figure 9. DNA detection by ImPyPyPy- γ -PyPyPyPy- β -Dp-TO.....	46
Figure 10. Titration of ImPyPyPy- γ -PyPyPyPy- β -Dp-TO with oligoduplex D1	47
Figure 11. Fluorescence of ImPyPyPy- γ -PyPyPyPy- β -Dp-TO.....	49
Figure 12. Fluorescence of ImPyPyPy- γ (TO)-PyPyPyPy- β -Dp.....	50
Figure 13. Fluorescence of ImPyPyPy- γ (TY)-PyPyPyPy- β -Dp	51
Figure 14. Fluorescence of ImPyPyPy- γ (TAB)-PyPyPyPy- β -Dp	52
Figure 15. DNase I footprint Titration of ImPyPyPy- γ (TY)-PyPyPyPy- β -Dp.....	54
Figure 16. "Molecular beacon" approach to DNA detection.	57
Figure 17. Polyamide detection of DNA using FRET.....	58
Figure 18. Synthesis of ImImPyPy- γ (BODIPY-Fl)-ImPyPyPy- β -Dp	60
Figure 19. Synthesis of ImImPyPy- γ (coumarin)-ImPyPyPy- β -Dp conjugate 19	61
Figure 20. Fluorescence of coumarin conjugate 19.....	63
Figure 21. Fluorescence BODIPY-Fl conjugate 18.....	64
Figure 22. Oligoduplexes used in these experiments	65
Figure 23. Energy transfer experiment.	66
Figure 24. Pyrene conjugates used in this study.	67
Figure 25. Computer model of pyrene polyamide binding modes.	70
Figure 26. Pyrene oligos used for pyrene excimer experiments.....	71
Figure 27. Fluorescence spectra of turn to turn pyrene system.....	73
Figure 28. Detection of DNA by coincidence detection	76
Figure 29. ^1H NMR of conjugate 14.	82

Chapter 3: Polyamide Conjugates for Affinity Purification of DNA

Figure 1. Polyamide affinity capture.....	88
Figure 2. Base pairs necessary to recognize unique sites	89
Figure 3. Biotinylating reagents.....	90
Figure 4. Synthesis of biotin linker derivative.....	91
Figure 5. Synthesis of biotin conjugates.	92
Figure 6. ImPy ₃ -β-ImPy ₃ -β biotin conjugates used in this study.....	93
Figure 7. Synthesis of His ₆ conjugate.....	94
Figure 8. Experimental capture procedures	95
Figure 9. Sequences of EcoRI/PvuII restriction fragments of pJT8 and pUC19.....	96
Figure 10. Representative capture data.	97
Figure 11. Capture experiment using polyamide beads	98
Figure 12. Model of a streptavidin-biotin-polyamide complex bound to DNA.....	100
Figure 13. DNase footprint titration of ImPyPyPy-γ-ImPyPyPy-β-Dp-XX-Biotin.....	101
Figure 14. DNase footprinting titration of ImPyPyPy-γ-ImPyPyPy-β-Dp-XX-Biotin..	102
Figure 15. Polyamide mediated affinity capture with ImPyPyPy-γ-ImPyPyPy-β-Ado ₃ His ₆ Gly.	104
Figure 16. Polyamide mediated affinity capture with preformed beads of ImPyPyPy-γ-ImPyPyPy-β-Ado ₃ His ₆ Gly.....	105
Figure 17. Capture and release of labeled restriction fragments.....	107
Figure 18. Quantitative DNase I footprinting of ImPy ₃ -γ-ImPy ₃ -β-Ado ₃ His ₆ Gly	108

Chapter 4: γ-Aminobutyric Acid Analogs for Polyamide Hairpin Formation

Figure 1. Improved binding of constrained hairpin polyamides.....	116
Figure 2. Constrained analogs of γ-aminobutyric acid	116

Figure 3. Comparison of α,γ and β,γ diaminobutyric acid turn residues	118
Figure 4. Synthesis of cyclopentyl amino acids.....	119
Figure 5. Synthesis of 2-Boc-amino-phenylacetic acid.....	119
Figure 6. Synthesis of <i>o</i> -Boc-aminomethylphenylacetic acid	120
Figure 7. Synthesis of <i>o</i> -Boc-aminomethylbenzoic acid	120
Figure 8. Synthesis of aminomethyl benzoic acid-pyrrole dimer.....	121
Figure 9. Synthesis of β,γ -diaminobutyric acid derivative, 33.....	122
Figure 10. Polyamides used in this study.....	123
Figure 11. DNase I footprinting of polyamides 37 (A) and 35 (B)	124
Figure 12. Computer model of constrained γ analog 39.	127
Figure 13. Summary of quantitative footprinting results.....	128
Figure 14. ^1H NMR of compound 33.....	137

Chapter 5: Recognition of G-Rich DNA Sequences

Figure 1. Restoration of high affinity binding through β -alanine linkers.....	142
Figure 2. Replacement of amide linkage	143
Figure 3. Polyamides designed to recognize 5'-WGGGGW-3' sequences.....	144
Figure 4. Retrosynthetic analysis of Boc protected alkyl diimidazole.	145
Figure 5. Failed halogenation of 1-methyl-4-nitroimidazole.....	146
Figure 6. Modified Hunsdiecker-Borodin Sonagashira reaction sequence.....	146
Figure 7. Synthesis of 1-methyl-4-nitro-2-acetylenoimidazole.....	147
Figure 8. Synthesis of 4 halo imidazoles.....	148
Figure 9. NOE difference spectrum of a mixture of 4 and 5 iodo imidazoles.....	149
Figure 10. Final Sonagashira coupling to give alkyne 26.....	150
Figure 11. Alternative route to give trimeric $\text{ImImCH}_2\text{CH}_2\text{Im-COOH}$	151

Chapter 6: Studies on the *isoG-isoC* Base Pair

Figure 1. The base pair between isoC and isoG	168
Figure 2. Consensus structure of the hammerhead ribozyme.....	170
Figure 3. Hammerhead kinetics experiment	172
Figure 4. Attempted EDTA autocleavage.....	174
Figure 5. Model of the hammerhead ribozyme	175
Figure 6. Synthesis of Ap-iGTP	176
Figure 7. Synthesis of d-isoG phosphoramidite	178
Figure 8. Synthesis of Me-isoC triphosphate.....	180
Figure 9. Design of transcription system.....	180
Figure 10. Transcription of d-isoG containing template.....	181
Figure 11. Proposed tautomeric equilibrium of isoguanosine	183
Figure 12. Potential base pairs of isoG tautomers	184

Appendix: Related Polyamide Studies

Figure 1. Structures of polyamides for the recognition of 5'-WGWCW-3' sequences ...	200
Figure 2. Sequence of the EcoRI/PvuII fragment of plasmid pSRC1	201
Figure 3. DNase I footprint titrations of compounds	202
Figure 4. DNase I footprinting of compounds 3 and 4	203
Figure 5. Binding isotherm for ImPyImPy- γ -PyPyPyPy- β -Dp.....	204
Figure 6. Synthesis of Fmoc-1-aminopropyl-4-Boc-amino-2-pyrrole carboxylic acid...	207
Figure 7. Polyamide glycoconjugate.....	207
Figure 8. Polyamide peptide conjugate.....	208
Figure 8. Synthesis of hydrophilic pyrrole derivative	208

Chapter 1:

Introduction to DNA Recognition

1.1 Biological Roles of Nucleic Acid Recognition

1.1.1 DNA Structure

The biological roles of DNA are intimately linked to its structure.¹ DNA is composed of a deoxyribose phosphate backbone with heterocyclic purine and pyrimidine nucleobases that participate in specific base pairing interactions. This specific Watson-Crick base pairing results in the now familiar double helical structure with adenine (A) paired with thymine (T) and guanine (G) paired with cytidine (C). This complementary base pairing, arising from specific hydrogen bonding interactions, is essential to DNA's role as an information storage molecule. As first noted by Watson and Crick,² a single strand can serve as a template for the synthesis of a complementary strand in the process of replication or transcription. The DNA double helix forms a wide and shallow major groove, and a deep and narrow minor groove. Each groove presents distinct hydrogen bonding faces at the edges of the base pairs, as shown in figure 2. Nature makes use of these faces for recognition in a wide variety of processes, including transcription, translation, replication, recombination, and viral integration.³

Recognition occurs at these positions through formation of hydrogen bonds and van der Waals contacts with the DNA bases. The major groove of an A-T base pair, for instance, displays a hydrogen bond acceptor in the N7 of adenine, a donor in the 6 amino group, an acceptor in the thymine carbonyl, and an opportunity for van der Waals

interactions with the thymine methyl group. This asymmetric arrangement of functionality allows for the discrimination of DNA sequences through formation of directional contacts in both the major and minor grooves.

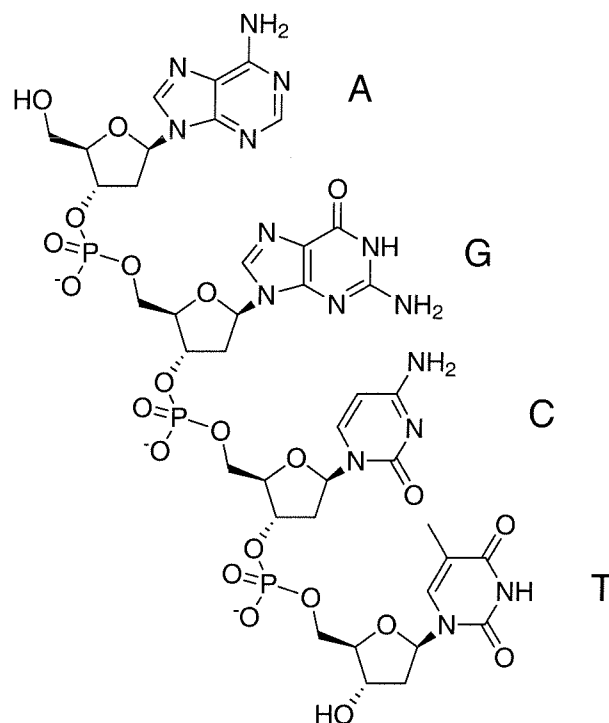


Figure 1. Chemical structure of DNA. The structure is drawn from 5' to 3', 5'-AGCT-3'.

This broad view of DNA structure is of course oversimplified. The canonical B-DNA structure is subject to many local, sequence dependent, conformational variations.^{4,5} The B-DNA structure, shown in figure 3, could well be described as a structure that averages a continuum of allowable conformational space. Detailed analysis of x-ray crystal structures and, more recently, of high resolution NMR^{6,7} structures have shown the flexibility and diversity attainable within the double helix. The broader range of conformations that DNA can adopt is even more evident in the comparison of the vastly distinct A and Z conformations.⁴

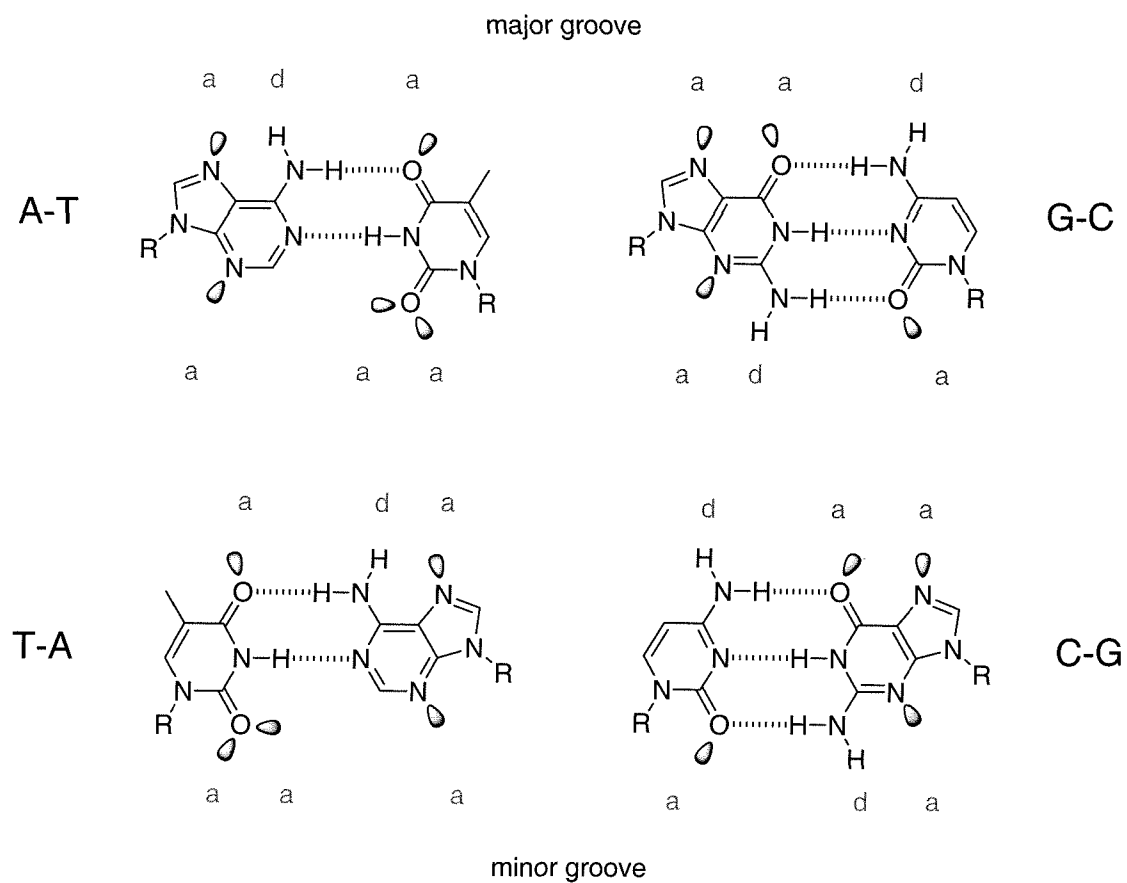


Figure 2. Hydrogen bonding in DNA base pairs. Lone pairs are shown with cartoon representations. Hydrogen bonds between base pairs are denoted by dashed lines, and the hydrogen bonding along the faces of the major and minor grooves are labeled with *a* or *d* for acceptor or donor.

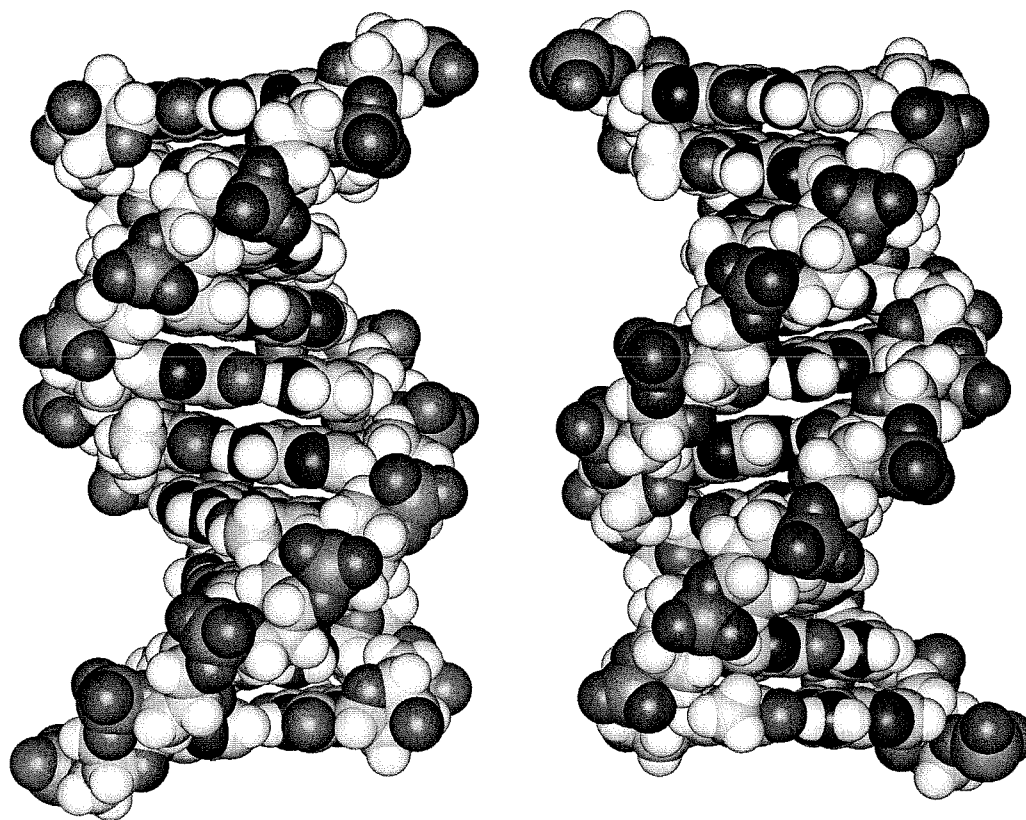


Figure 3. Model of B-form DNA. Spacefilling representation of DNA was generated in InsightII with normalized B DNA parameters.

1.1.2 Transcription Regulation

Expression of the information contained in DNA is carried out through the processes of transcription and translation, that is the generation of a complementary copy of the DNA in the form of RNA and synthesis of the encoded protein from the RNA template.



Figure 4. Central dogma of molecular biology.

The regions upstream of a particular gene sequence contain binding sites for transcription factors that can, through concerted action, either enhance or repress the level of transcription.

Specifically, eukaryotic genes encoding most proteins are transcribed by RNA polymerase II (Pol II). Most Pol II promoters require the presence of a set of general transcription factors, including the TATA binding protein (TBP), and several TBP associated factors (TAFs).⁸ The assembly of these factors is sufficient for basal levels of transcription, but their roles *in vivo* are still in debate.⁹

Activated transcription occurs in response to developmental or environmental signals requiring the increased expression of a specific gene. Specific transcription factors bind to the promoter region or farther upstream in the enhancer, and lead to increased levels of transcription. Several mechanisms can explain the activation, including recruitment of the basal transcription complex,¹⁰ disruption of chromatin structure,¹¹ and stabilization of the reinitiation complex.¹² The promoter region can also be tuned to a variety of inputs, so different combinations of transcription factors can effect transcriptional activation. This has been shown to be especially true during development.¹³

A major goal in the design of sequence-specific DNA binding molecules is the specific modulation of this process of transcription, through the generation of synthetic activators and/or repressors. Recent success in the Dervan group has focused on the HIV-1 promoter. In the interest of illustrating the complexity of DNA-protein interactions in a promoter region, figure 5 shows a computer generated model of the HIV-1 promoter based on crystal structures of the proteins or homologous proteins bound to the DNA. From this crude model, the role of DNA binding transcription factors in determining a stereospecific promoter complex is quite obvious. Not only do factors need to bind, but location and orientation within the promoter complex are essential. DNA bending architectural factors can induce protein-protein contacts between proteins located on distal DNA binding sites.

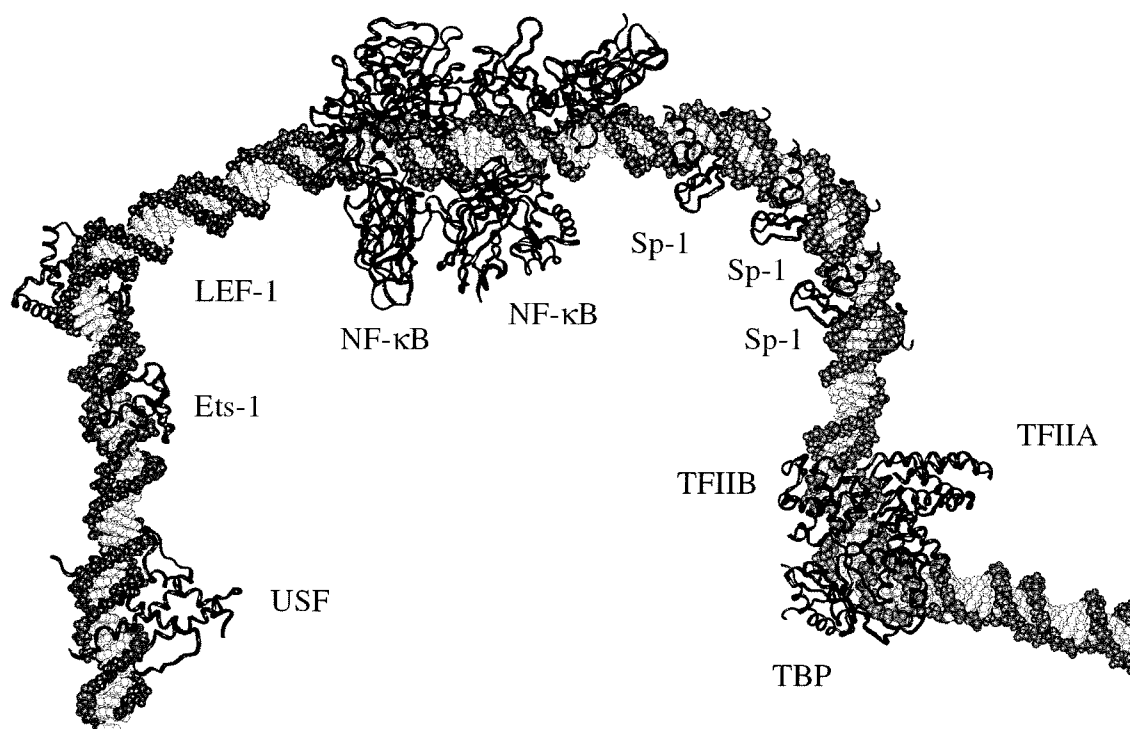


Figure 5. Model of the HIV-1 promoter. Model was generated using InsightII according to the HIV-1 promoter sequence and binding site information.¹⁴ Individual crystal or NMR structures of USF,¹⁵ PU.1¹⁶ (in place of Ets-1), LEF-1,¹⁷ NF-κB (p50 homodimer¹⁸ in place of the p50/p65 heterodimer), Zif268¹⁹ (in place of Sp1), TFIIA,²⁰ TFIIB,²¹ and TBP^{20,21} were joined with appropriate spacing according to the promoter sequence. Bending at the Sp1 site is substantiated by biochemical evidence for the Sp1 family.²²

1.2 Protein and Macromolecule DNA Interactions

1.2.1 DNA Binding Motifs

The transcription factors necessary for activated transcription, as well as many other proteins, must recognize specific DNA sequences in order to function. Several protein motifs have evolved that efficiently bind DNA.^{8,23-25} Most common motifs place an alpha-helix in the major groove of the DNA, with the amino acid side chains providing base-specific contacts. Examples of some common DNA binding motifs are shown in figure 6, including zinc fingers, leucine zippers, helix loop helix, and homeodomain classes. There

are several other motifs of transcription factors including the ETS domain, Gal4 zinc binding domain, and steroid receptors, as well as distinct permutations of each class.

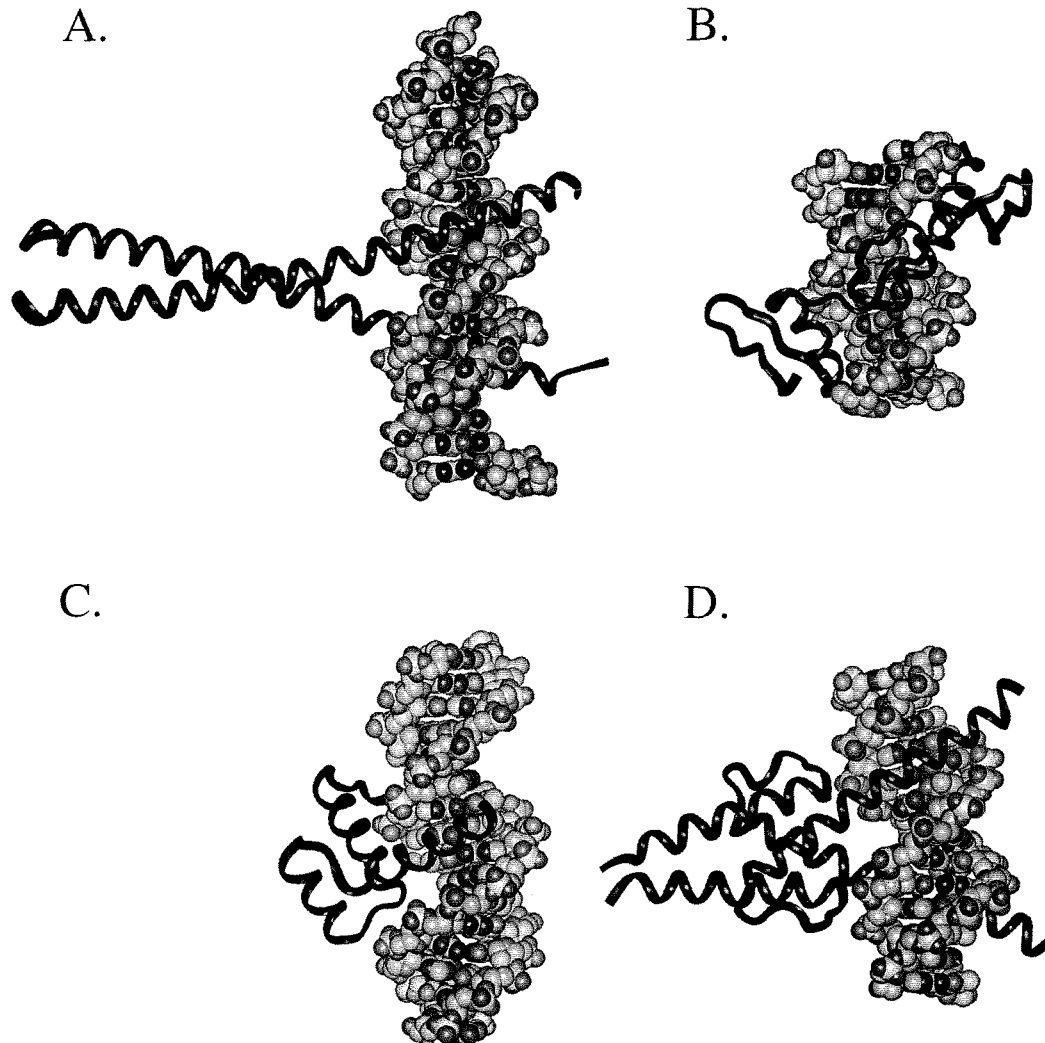


Figure 6. Common transcription factor DNA binding motifs. A) GCN4, bZIP or leucine zipper;²⁶ B) Zif268,¹⁹ zinc finger; C) Engrailed,^{27,28} homeodomainD); MyoD,²⁹ Helix-Loop-Helix (HLH).

While the majority of proteins discussed above make their contacts in the major groove of DNA with an α -helical region of the protein, several distinct protein types have been discovered that bind predominantly in the minor groove of DNA. These include TBP^{30,31}

and several architectural transcription factors, including the HMG and Lef-1 proteins. In addition, several proteins make specific contacts in both the major and minor grooves. Two such proteins bind the HIV-1 promoter: USF and Ets-1.

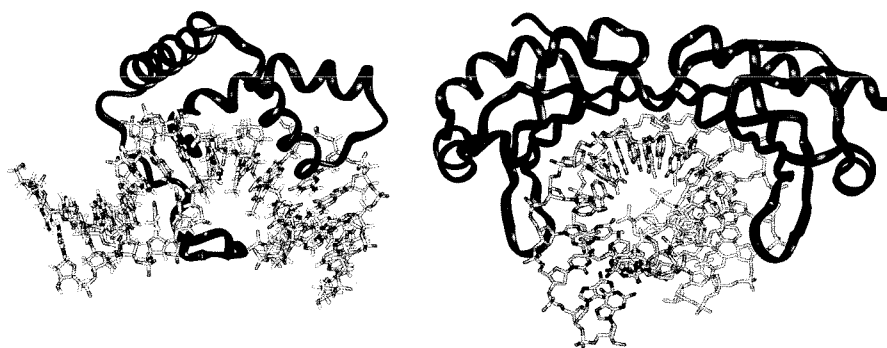


Figure 7. Minor groove binding transcription factors. On the left is the NMR structure of LEF-1. On the right is a crystal structure of TATA Binding Protein (TBP).^{30,31}

1.2.2 Design of Sequence-specific DNA Binding Proteins

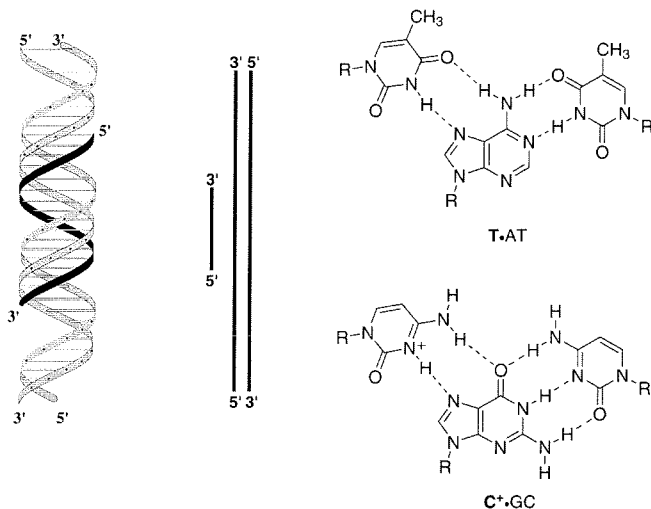
An attractive approach to the recognition of DNA is the design of sequence-specific DNA binding proteins. A number of factors influence the generality of sequence-specific protein binding to DNA. First, no simple code links amino acid residues in these binding motifs to DNA bases in recognition.^{23,24} Second, for certain applications, proteins are not appropriate. Lastly, proteins are large, usually cell impermeable and expensive to produce. With these cautions aside, several successes have been reported in the design of sequence-specific DNA binding proteins. Two groups have successfully pursued the approach of tethering two unrelated DNA binding regions to make a hybrid recognition motif.³²⁻³⁵ This rational design approach is effective at generating proteins with larger binding sites but cannot rapidly lead to molecules targeting large numbers of sites. The second, and

possibly more general, approach has been pursued both by Pabo and coworkers³⁶ and Klug and coworkers.^{37,38} In the zinc finger motif described above, a single zinc finger recognizes 3 base pairs of double helical DNA. Through a combination of rational design and affinity selection using phage display libraries, artificial zinc finger proteins binding to a number of diverse sequences have been generated. The ability of the phage display system to provide wide diversity has allowed for the rapid selection of proteins targeting a large number of sequences.³⁸ While this technique is quite powerful, there are significant disadvantages of the use of proteins for DNA recognition rather than small molecules, particularly for *in vivo* applications. However, a novel transcription factor expressed within a cell can modulate transcription in a sequence-specific fashion, opening the possibility for the use of such proteins in gene therapy approaches.³⁷

1.2.3 DNA Recognition by Triple Helix Formation

Another powerful approach to the sequence specific recognition of DNA is oligonucleotide directed triple helix formation.^{39,40} In triple helical nucleic acids, a third strand of nucleic acid binds in the major groove of DNA, with the bases forming specific hydrogen bonds to the faces of the major groove. These interactions are shown in figure 8. Two motifs of triple helix formation have been elaborated, the pyrimidine⁴⁰ and purine motifs, named for the predominant type of base in the third strand. Significant limitations exist in the pairing rules for triple helix formation. Specifically, targets are limited to sites with large homopurine runs. Significant effort has been placed on expanding the recognition code of the triple helix approach through the design of non-natural bases. While a number of these have shown some promise, the overall success has been quite poor.

A. Pyrimidine Motif



B. Purine Motif

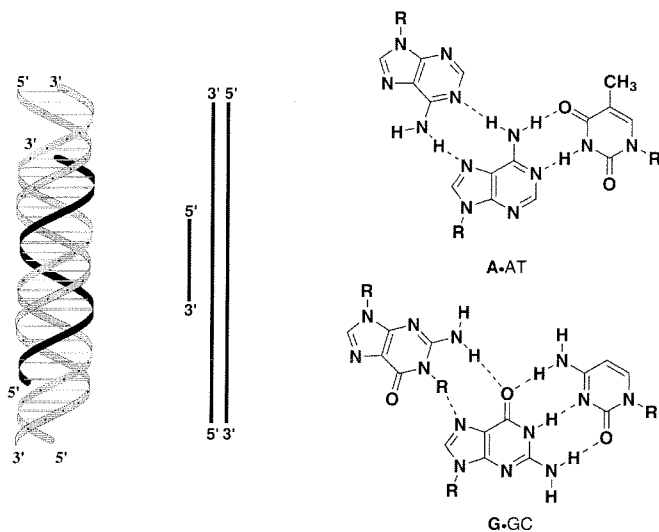


Figure 8. Base triplet structures for pyrimidine and purine motif triple helices.

In spite of these shortcomings, triple helix forming oligonucleotides are powerful in their ability to recognize DNA. Third strand oligonucleotides can bind with extremely high affinity, depending on the length of the sequence, and with unparalleled sequence specificity. Single base mismatches in a binding site for a triple helix forming oligonucleotide result in two to three order of magnitude decreases in affinity.^{41,42} This

affinity and specificity has been applied to recognize a single site within a human chromosome,⁴³ to regulate gene transcription,^{44,45} and to inhibit viral replication.^{46,47}

The promise of these triple helix forming oligonucleotides for use as modulators of biological processes *in vivo* has been more significantly limited by the limited cell permeation of these molecules.⁴⁸ In spite of the large scientific community pursuing triple helixes as potential therapeutics, there have been remarkably few successes. Delivery of oligonucleotides to cells remains a major challenge in this field. Significantly, the antisense field which relies on similar compounds has also shown little clinical success, with the exception of Isis Pharmaceutical's recently approved Vitravene™ (fomivirsen sodium injectable) for the treatment of cytomegalovirus (CMV) retinitis in patients with AIDS. The second major drawback to triple helix formation for the *in vivo* control of gene expression is avoiding biochemical degradation or compartmentalization of the oligonucleotides. As the active agents in triple helix formation are oligonucleotides with various modifications, they are often susceptible to cellular machinery that can bind to or degrade the reagent.

1.2.4 DNA Recognition by Polyamide Nucleic Acid (PNA)

A similar approach to the triple helix approach is recognition by polyamide nucleic acids. PNA is a DNA-like oligomeric material in which nucleobases are attached through an aliphatic amide structure.^{49,50} PNA can interact with DNA with high affinity and specificity. Unlike DNA, PNA will readily form D-loop structures with DNA duplexes.⁵¹ In such a structure, recognition occurs by Watson-Crick base pairing to one strand while the complementary strand is displaced. This mode of recognition overcomes the sequence limitations of the triple helix approach. As the repeating unit of PNA is uncharged, it was hoped that it would be more suitable for *in vivo* applications than was DNA. Unfortunately, cells have proved to be relatively impermeable to PNA. The high affinity

and specificity of PNA for various nucleic acid targets has provided numerous applications of PNA in hybridizations and other molecular biology manipulations.⁵²

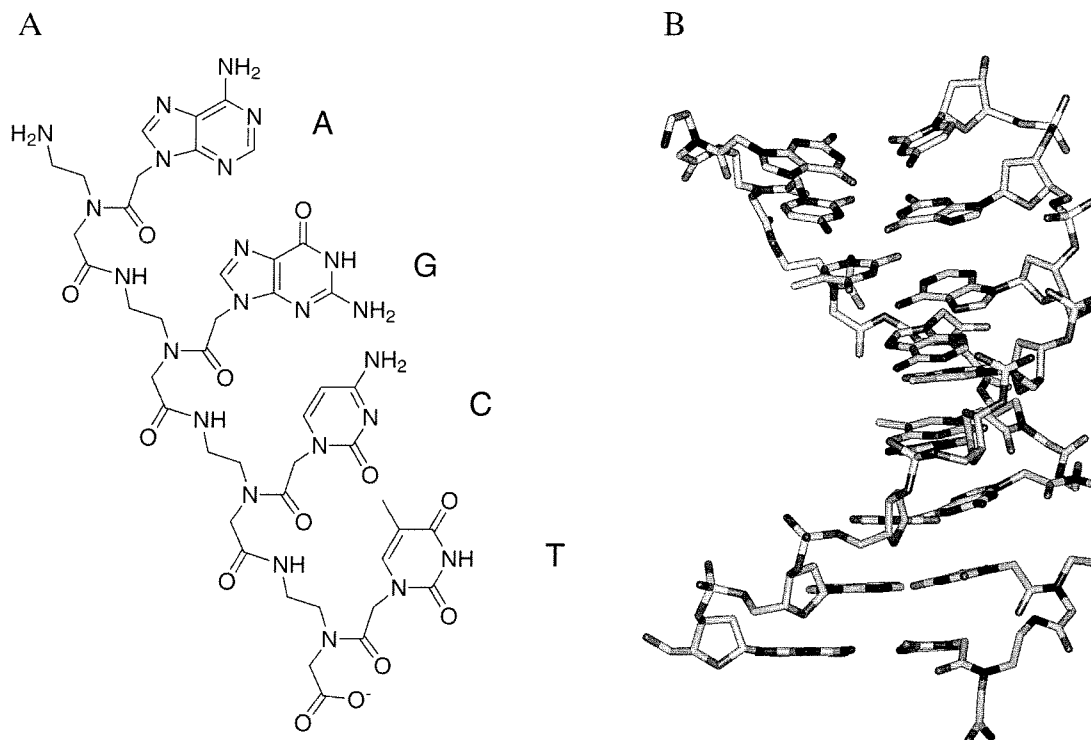


Figure 9. PNA Structure. A) Chemical structure of PNA B) Structure of a PNA-DNA duplex.⁵³

1.3 Small Molecule-DNA Interactions

In contrast to protein recognition of DNA, small molecules provide a distinct framework for the design of sequence-specific DNA binding reagents. The clinical usage of some DNA binding compounds has driven research in this area. Small molecules can interact with DNA through distinct binding modes, intercalation or groove binding (through the major or minor grooves), or in combinations of both.

1.3.1 Intercalation

A common mode of natural product interaction with double stranded DNA is through intercalation, or stacking within the base stack. Intercalation results in the

separation of the DNA heterocyclic bases, effectively unwinding the helix. Intercalators can interact with DNA with quite high affinity, the result of desolvation and aromatic stacking (polarization) interactions. A number of intercalating molecules are shown in figure 10. All possess a reasonably large aromatic surface that is buried in the DNA base stack. Affinity is balanced, however, by the large energetic penalty of creating the distorted intercalation site in the DNA.⁵⁴

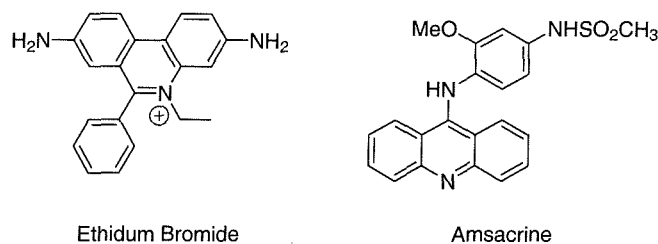


Figure 10. Intercalative DNA binding compounds.

A number of more complex intercalators have been designed and synthesized. One common class is the bis-intercalators. This class of compounds show large binding enhancements through the increased interaction of the two halves with the DNA. Several of this class of compounds come close to achieving an additive energetic contribution of the two halves. Early work in the Dervan group led to the design and synthesis of an ethidium dimer.⁵⁵ More recently, cyanine dye dimers, such as the thiazole orange dimer (TOTO), have proved to be quite successful as reagents for detecting DNA with very high affinity.^{56,57} Chaires and coworkers have designed and synthesized another dimer, WP631, that binds with high affinity that is unexpectedly enthalpically driven.^{58,59}

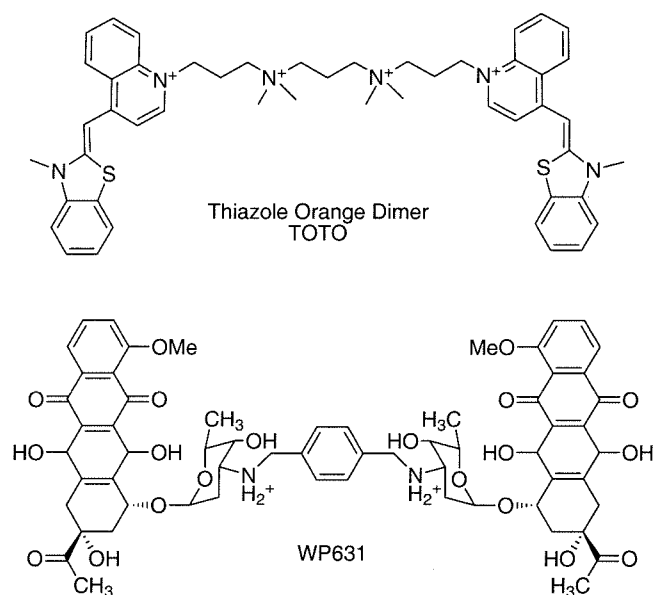


Figure 11. DNA binding bis-intercalators: Thiazole orange dimer^{56,57} and WP631.^{58,59}

A prominent feature of some intercalative DNA binders is that they are fluorescent, and show increased fluorescence in the presence of DNA. In the case of ethidium bromide, the most commonly used double stranded DNA (dsDNA) dye, this 40-fold enhancement results mainly from solvent exclusion. Solvent exclusion effects increased fluorescence by disfavoring the thermal release of energy to neighboring solvent. Fluorescent enhancements of cyanine dyes are even greater,⁵⁶ and will be discussed in more depth in Chapter 2.

1.3.2 Groove Binding

The second major class of DNA binding agents are the groove binders. Groove binders are characterized by extended planar aromatic structures that interact with the narrow deep minor groove of DNA.^{60,61} Classic examples of minor groove binding compounds include distamycin, Hoechst 33258, CC-1065, and SN 6999, all shown in figure 12. The energetics of binding have been extensively characterized for some of these

compounds. For Hoechst 33258, for example, binding energy correlates with the extent of hydrophobic surface buried in the groove, not with the extent of hydrogen bonding.⁶²

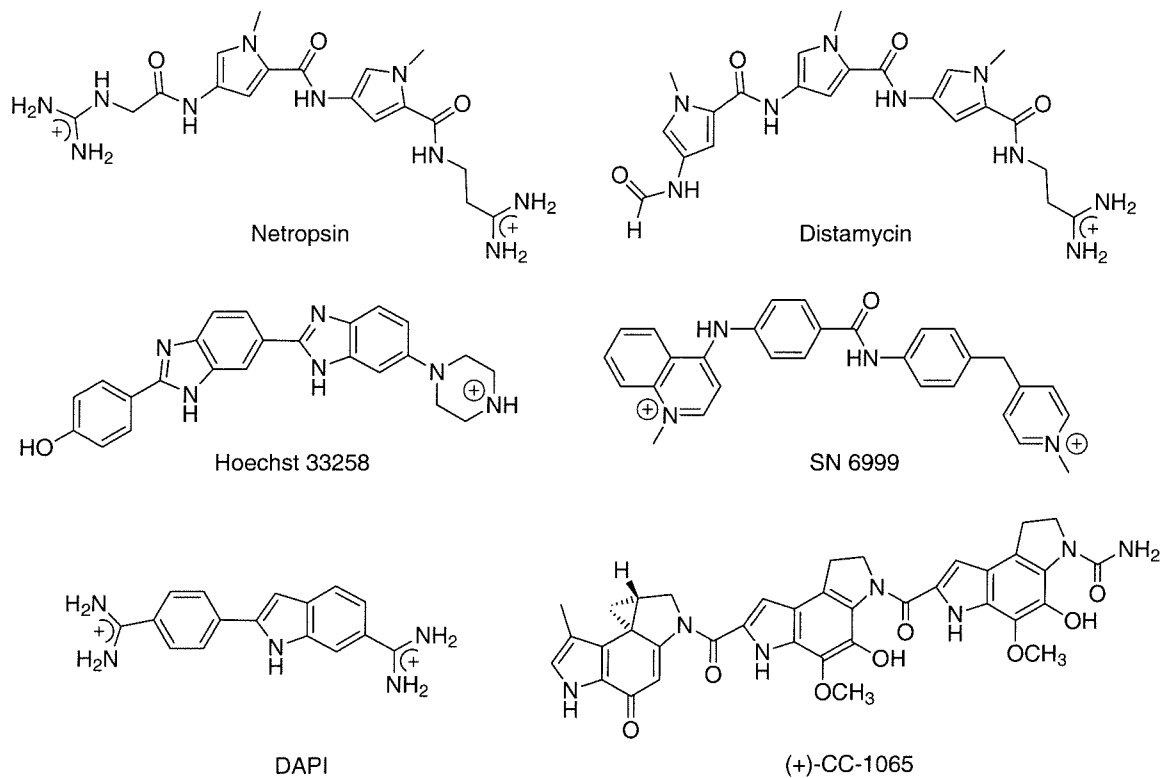


Figure 12. Minor groove binding compounds.

Much like the intercalators, these compounds share a common feature of an extended aromatic surface. Unlike the intercalators, minor groove binders achieve binding by nestling in the deep narrow minor groove. This results in substantial desolvation and increased van der Waals contacts with the walls of the groove. In most cases of either natural products or designed/analogs, these compounds also form hydrogen bonds to the floor of the minor groove. These hydrogen bonding interactions are essential for the specificity of the compounds for particular DNA sequences; however, as a hydrogen bond to solvent must be disrupted prior to binding, their energetic contributions may not be extremely significantly positive.

1.3.3 Simultaneous Groove Binding and Intercalation

Finally, many natural and synthetic products interact with DNA in a combination of these binding modes. The anthracycline family of antibiotics, for example, intercalate an aromatic quinone group into the DNA and bind the groove with bulky carbohydrate arms. The chemotherapeutic agent Actinomycin D interacts similarly, but with cyclic peptides binding to the minor groove.⁶¹

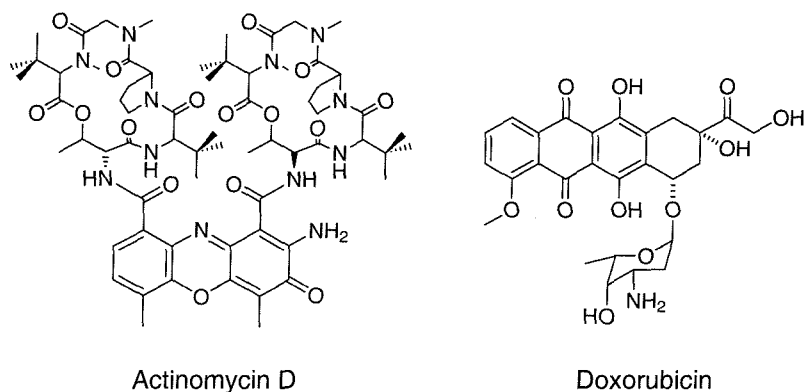


Figure 13. Ligands that interact through intercalation and groove binding.

Barton and coworkers have also contributed significantly to the design of sequence specific DNA intercalators.⁶³ These bulky transition metal complexes interact predominantly with the major groove of DNA,⁶⁴ though there is some debate in the literature.^{65,66} The use of the metal complex to guide functional groups to the faces of the major groove makes it quite unique in the DNA recognition field, as there are few small molecules that interact with the DNA in the major groove.

1.3.3 Distamycin Analogs

The modular nature of the repeating amide structure of the natural products distamycin and netropsin has generated a large synthetic effort in producing analogs.⁶⁷ Much early work in this field was predicated on the 1:1 binding model from crystal

structures of Rich and coworkers. This structure is shown in figure 14. Distamycin is tightly bound to the DNA, with large van der Waals contacts with the walls of the minor groove. Distamycin preferably binds to A/T rich DNA sequences, and the amides form hydrogen bonds to the DNA bases in the minor groove. The general distamycin structure is remarkably tolerant of small changes. The five membered heterocyclic pyrrole rings have been replaced with imidazoles, furans, thiazoles, and other heterocycles. Indeed, synthetic molecules with similar shapes based on six membered rings also bind.⁶⁷ In spite of the broad interests in these classes of compounds, few efforts have focused on understanding and utilizing the sequence specificity of these compounds.

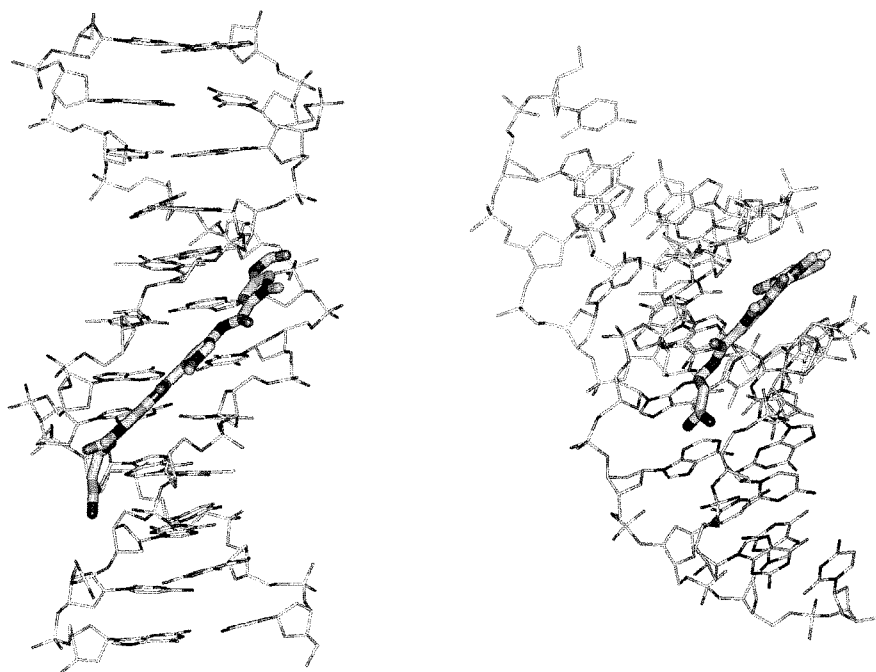


Figure 14. Structure of 1:1 complex of distamycin with DNA.⁶⁸

1.3.4 Hybrid Binding Molecules

A large number of compounds have been designed that combine components of two DNA binding ligands. These hybrid molecules have distinct properties, and sometimes

novel binding modes. Distamycin has been conjugated to a number of other moieties, including the intercalating compounds, acridine and psoralen, and the fluorescent compound pyrene. This area has also been thoroughly reviewed.⁶⁹

1.3.5 Functional Distamycin Conjugates

A second focus of research in the area of DNA recognition is coupling recognition with chemical reactivity or other function. A wide variety of conjugates have been prepared with DNA alkylators and other cleavage reagents. Such DNA cleaving compounds could find utility as artificial restriction endonucleases. Longer term research aims to develop more selective chemotherapeutic agents, particularly alkylators that target specific disease genes. This area of research is the subject of a recent review.⁶⁷

1.4 Minor Groove Binding Polyamides

The term polyamide will be used throughout this thesis to refer to oligomers of imidazole (Im), pyrrole (Py), and hydroxypyrrole (Hp) carboxamides with appropriately spaced aliphatic amino acids, β -alanine (β) and γ -aminobutyric acid (γ). These polyamide compounds are distamycin derivatives, but differ in their preferred 2:1 binding motif, their high affinity binding, and in their sequence-specific binding. For the purposes of discussion throughout this thesis, several definitions must be made. Polyamide sequences are always written in an N-terminal to C-terminal sense. All compounds containing an N-terminal imidazole contain 1-methyl-2-imidazole carboxylate rather than the internal 4-amino imidazole. Most of the polyamides discussed herein contain a *N,N*-dimethylaminopropylamide (Dp) positively charged tail. A convenient cartoon representation of these polyamides has been developed in the Dervan group. Imidazoles, both terminal and internal, are represented by filled circles; pyrroles by open circles; hydroxypyrroles by circles surround the letter H; β -alanine by a diamond; γ -aminobutyric

acid (γ) by a larger half circle; and the dimethylamino propylamide tail is represented by a half circle surrounding a “+” sign. This type of cartoon is illustrated in figure 15.

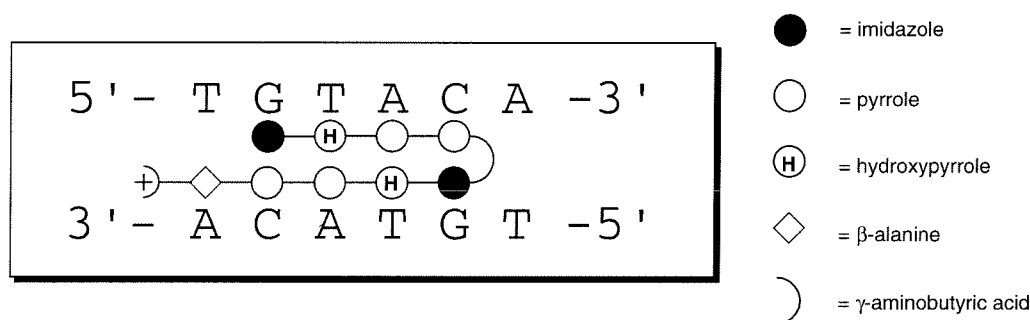


Figure 15. Schematic polyamide representation. The polyamide shown in this example is ImHpPyPy- γ -ImHpPyPy- β -Dp and targets a 5'-WGTACW-3' sequence in which W indicates either an A or T in the binding site.

1.4.1 Dimeric binding

The early crystal structures solved by the Rich and Dickerson groups provided an incomplete model of the structural diverse binding modes distamycin could achieve.⁶⁸ Observation of distamycin bound in the minor groove of DNA as a side by side antiparallel dimer by NMR provided a clear model for interpreting footprinting results obtained with early imidazole containing distamycin derivatives.⁷⁰ The landmark NMR structure determined by Wemmer and coworkers⁷⁰ provided a framework for the interpretation of the work of Wade and Dervan on the sequence specificity of a polyamide of the sequence ImPyPy-Dp.^{71,72} Placement of an *N*-terminal imidazole or pyridine residue in place of the *N*-terminal pyrrole carboxamide of distamycin (abbreviated ImPyPy-Dp) resulted in binding to 5'-WGWCW-3', where W indicates either A or T.⁷³⁻⁷⁵ In more recent work, this 2:1 binding mode has been confirmed in crystal structures of distamycin and of polyamide molecules. Analysis of this sequence in the context of the 2:1 motif provided the

key necessary to construct the pairing rules discussed below, namely that imidazole residues specifically recognize guanine bases through specific hydrogen bonding interactions. The ability of the minor groove to accommodate either a single compound or pair of compounds illustrates the conformational flexibility of DNA.

1.4.2 Pairing Rules for Polyamide Binding

With the preference of imidazole residues for guanine residues established, the Dervan group has defined the pairing rules for polyamide binding containing imidazole and pyrrole residues. Remarkably, an imidazole residue recognizes the guanine amino group only on one side of the groove. An imidazole-pyrrole pair has specificity for G-C and not C-G. Pyrrole-pyrrole pairs are degenerate for A-T and T-A base pairs within a range of 1.5 kcal/mol.⁷⁶ White *et al.* studied a hydroxypyrrole derivative that makes specific hydrogen bonds to T residues, completing the pairing code.⁷⁷ The overall pairing rules are summarized in figure 16. In short, imidazole-pyrrole pairs recognize G-C base pairs, pyrrole-imidazole pairs recognize C-G base pairs, hydroxypyrrole-pyrrole pairs recognize T-A base pairs, pyrrole-hydroxypyrrole pairs recognize A-T base pairs, and pyrrole-pyrrole base pairs are degenerate for A-T and T-A. In some cases it is necessary to replace a pyrrole with an aliphatic amino acid, especially β -alanine.⁷⁸ Replacement of the rigid pyrrole unit with β -alanine has been shown to increase binding energies, perhaps through relief of steric constraints imposed by mismatched curvature or register of the polyamide relative to that of the DNA. Using β -alanine spacers, compounds have been prepared that bind sequence of up to 11 base pairs as completely overlapped compounds, and up to 16 base pairs in slipped binding modes.^{79,80}

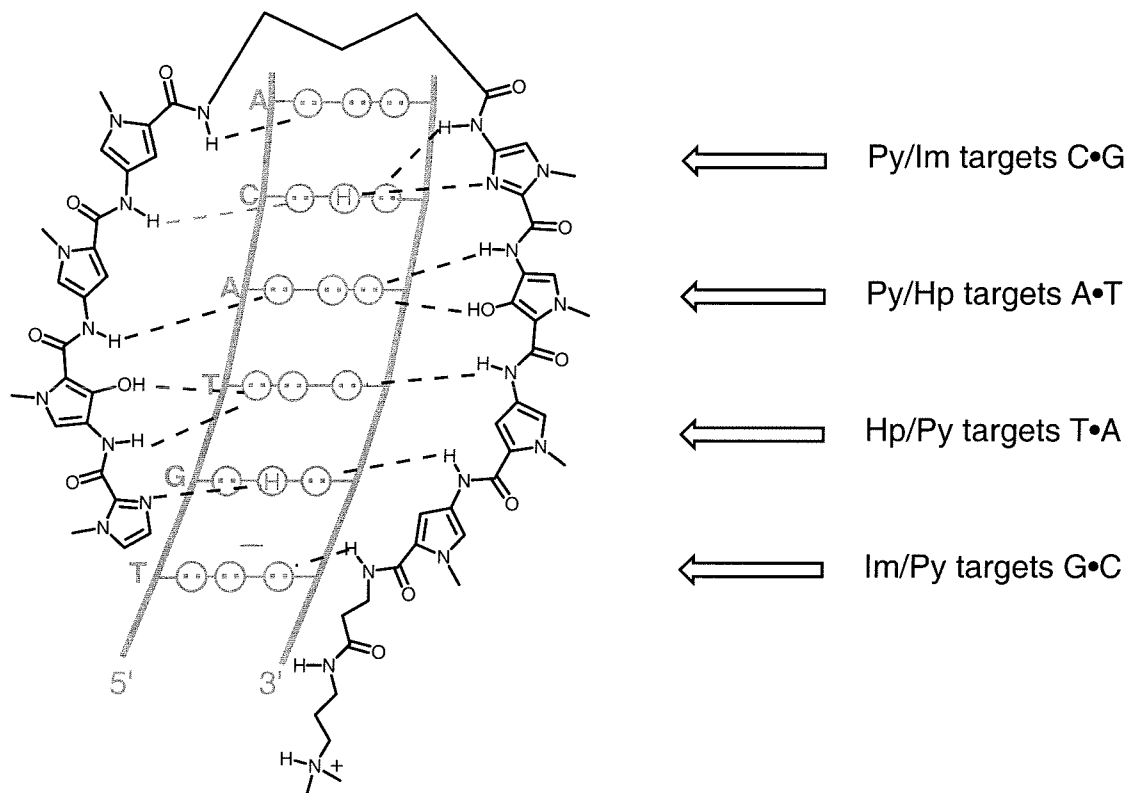


Figure 16. Pairing rules for DNA recognition by polyamides.

More recently, these pairing rules have been confirmed, and greater structural detail revealed through the solution of x-ray crystal structures of 2:1 polyamide complexes with DNA in a collaboration with the Rees group at Caltech. Structures have been solved for the polyamides ImPyPyPy- β -Dp, ImImPyPy- β -Dp,⁸¹ and ImHpPyPy- β -Dp.⁸²

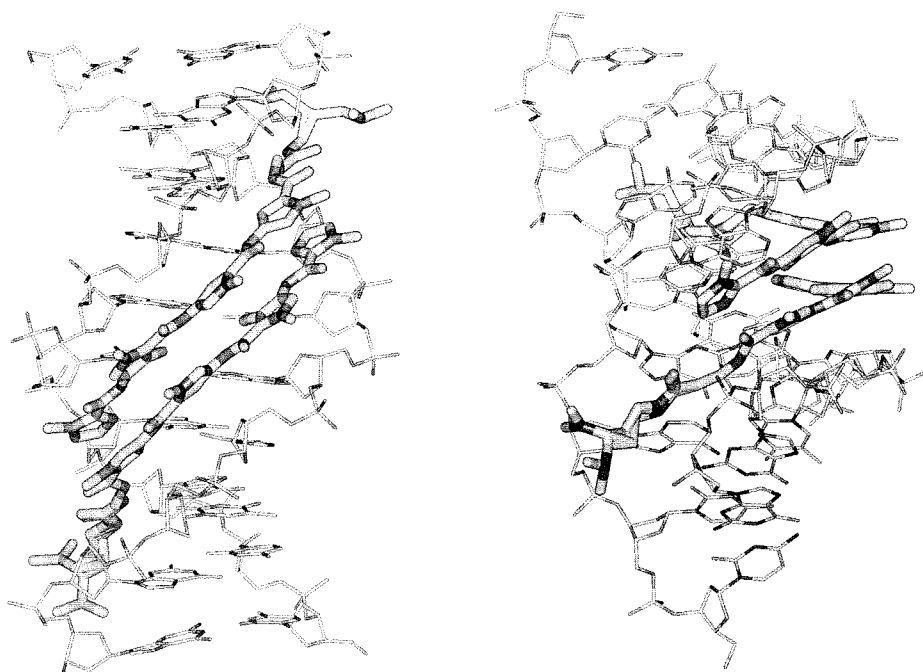


Figure 17. X-ray crystal structure of ImHpPyPy- β -Dp.⁸²

1.4.3 Polyamide Motifs

Discovery of the 2:1 polyamide binding motif inspired the development of a new class of polyamides. Linkage of side by side polyamides through pyrrole aromatic nitrogen (an H-pin polyamide)⁸³ or through an N terminal to C terminal linkage (a hairpin)⁸⁴ results in generally higher affinity compounds than the unlinked pairs. Further rigidification through cyclization of both termini of the polyamide results in further energetic gain,⁸⁵ with somewhat reduced specificity, although this is still under consideration.

Hairpin molecules have proved to be most useful in terms of a combination of synthetic ease, affinity and specificity.⁴¹ Indeed, a hairpin polyamide, ImPyPy- γ -PyPyPy- β -Dp had affinity 400 fold higher than the parent unlinked polyamides for the same recognition site, with the advantage of controlling slippage and homodimerization in the binding motif.⁸⁴ As the length of the recognition site increases, β -alanine residues must be

used to restore high affinity binding. The 2- β -2- γ -2- β -2- β -Dp motif, first investigated by J. Trauger, has proved to be an exceptionally versatile scaffold for the design of polyamides recognizing seven base pair binding sites.

For longer recognition sites, unlinked dimers have proved useful. One advantage of such a system is that such compounds effectively double the binding site available for a given molecular weight. A number of motifs have been established for this extended binding, including the 3- β -3,⁸⁶ 4- β -4,⁷⁹ and 2- β -2- β -2- β -2⁸⁰ motifs. The numbers correspond to the number of aromatic amino acid residues placed according to the aforementioned pairing rules. All motifs share the common attribute of using β -alanine residues to allow high affinity interaction. Early experiments by S. Youngquist with distamycin analogs in a similar motif clearly foreshadowed the importance of β -alanine residues in the recognition of lengthy DNA sequences.⁸⁷ More recently, Trauger *et al.* showed that a compound of the structure ImPy- β -ImPy- β -ImPy- β -PyPy- β -Dp binds a sixteen base pair sequence in the HIV promoter with high affinity and excellent specificity.⁸⁰

1.4.4 Synthesis of Polyamides

A major advance in the synthesis of polyamide molecules was the development of solid phase synthesis methodologies by E. Baird.⁸⁸ Using standard *tert*-butoxycarbamate (Boc), or more recently, fluorenylmethylcarbamate (Fmoc) protecting groups, and solid phase synthetic methodologies, polyamide molecules can be prepared in days, in reasonable scales (1-100mg routinely), and in relatively high purity. All polyamide compounds in this thesis were prepared using these solid phase synthesis techniques, with some slight modifications, as discussed within. The ability to synthesize compounds rapidly has contributed greatly to the understanding of DNA recognition. Even before the

adoption of solid phase synthesis, synthesis of polyamides was easily an order of magnitude faster than comparable DNA binding drugs such as CC-1065. With the advent of the solid phase approach, increased numbers of compounds have been assayed for DNA binding, resulting in an efficient feedback cycle.

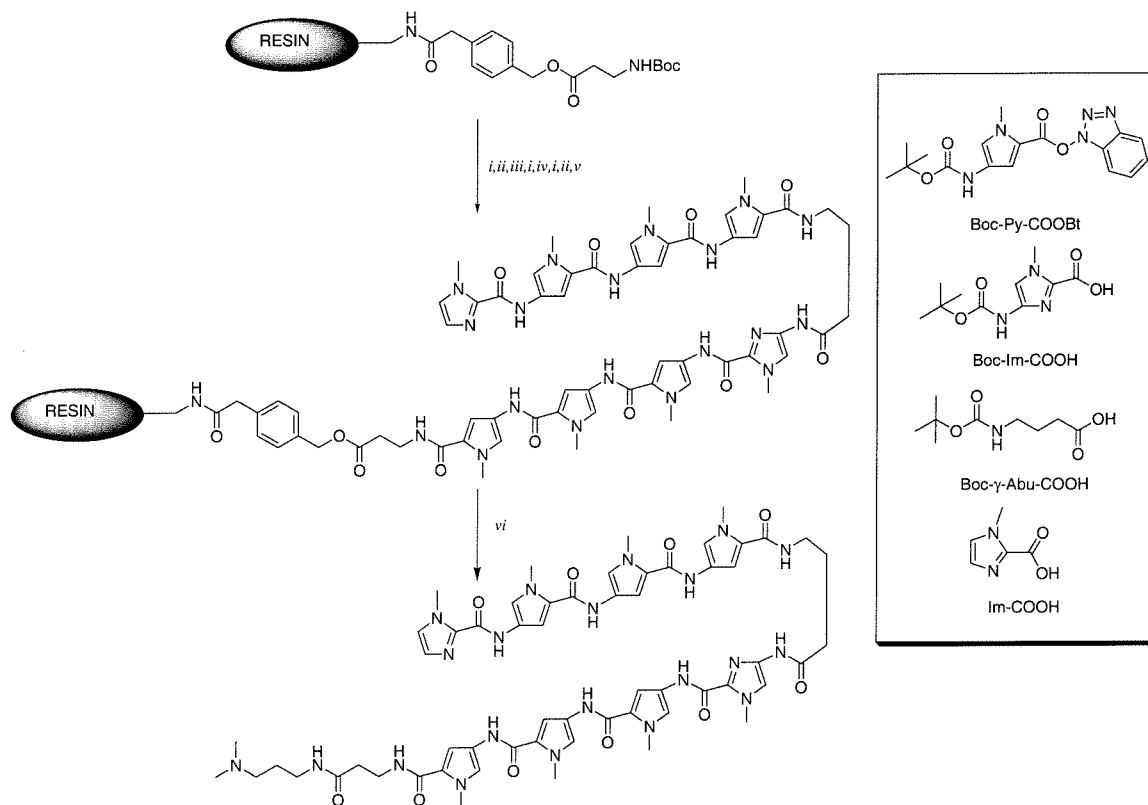


Figure 18. Solid phase synthesis of polyamides.⁸⁸ i) trifluoroacetic acid, dichloromethane, thiophenol; ii) Boc-Pyr-COOH; trifluoroacetic acid; 3 times; iii) Boc-Im-COOH, HBTU, DIEA; iv) Boc-γ-Abu-COOH, HBTU, DIEA; v) Im-COOH, HBTU, DIEA; vi) *N,N*-dimethylaminopropylamine, 55°C, 12 hours.

1.4.5 Polyamide Conjugates

The high affinity and sequence specificity of polyamide molecules has been put to use as a means for directing various reactive groups to the DNA in a sequence specific fashion. Conjugates have now been prepared with oligonucleotides,^{89,90} peptides, peptides and oligonucleotides, intercalators, alkylators, and various DNA reactive compounds. In many cases conjugation leads to little change in the specificity or affinity of

the polyamide for DNA, while in some cases larger effects are seen. Several conjugates of polyamides with fluorescent dyes, biotin, or histidine tags are discussed within this thesis. Much of the work in this area is ongoing.

1.5 This Work

This thesis describes two applications of minor groove binding polyamides. The first is the use of polyamides as fluorescent probes of DNA. In Chapter 2, efforts to design, synthesize, and characterize fluorescent conjugates of minor groove binding polyamides are described. In Chapter 3, polyamides were used as reagents for purification of double stranded DNA with the goal of attaining sequence specific purification reagents.

Chapters 4 and 5 present efforts towards the improvement of the affinity and specificity of polyamide molecules through synthetic modifications. Specifically, Chapter 4 describes the synthesis of polyamides that contain rigid analogs of γ -aminobutyric acid designed to decrease the entropic penalty incurred in the 2:1 motif. In addition, a charged β,γ -diaminobutyric acid was synthesized that could make favorable hydrogen bond contacts with the floor of the minor groove and serve as a point of attachment for other functional moieties. Chapter 5 summarizes synthetic efforts towards the recognition of G rich DNA sequences through the replacement of an imidazole-imidazole amide linkage with an alkyl linkage. Appendix 1 provides an appendix of related synthetic work not developed fully, but which could be useful for further work in the polyamide area.

Chapter 6 summarizes early efforts studying the non-natural base pair between isoguanosine and isocytosine. Deoxyisoguanosine was synthesized suitably protected for solid phase synthesis, and its ability to direct transcription assayed. This non-natural base pair was also exploited in the study of a hammerhead ribozyme.

1.6 References

- (1) Saenger, W. *Principles of nucleic acid structure*; Springer-Verlag: New York, 1985.
- (2) Watson, J.; Crick, F. *Nature* **1953**, *171*, 964-967.
- (3) Blackburn, G. M.; Gait, M. J. *Nucleic acids in chemistry and biology*; Oxford University Press: Oxford; New York, 1996.
- (4) Dickerson, R. E. *Meth. Enz.* **1992**, *211*, 67-111.
- (5) Joshua-Tor, L.; Sussman, J. L. *Curr. Op. Struct. Biol.* **1993**, *3*, 323-335.
- (6) Feigon, J.; Sklenar, V.; Wang, E.; Gilbert, D. E.; Macaya, R. F.; Schultze, P. *Meth. Enz.* **1992**, *211*, 235-253.
- (7) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Res. Spect.* **1998**, *32*, 287-387.
- (8) Patikoglou, G.; Burley, S. K. *Ann. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 289-325.
- (9) Moqtaderi, Z.; Bai, Y.; Poon, D.; Weil, P. A.; Struhl, K. *Nature* **1996**, *383*, 188-191.
- (10) Klein, C.; Struhl, K. *Science* **1994**, *266*, 280-282.
- (11) Kadonaga, J. T. *Cell* **1998**, *92*, 307-313.
- (12) Ho, S. N.; Biggar, S. R.; Spencer, D. M.; Schreiber, S. L.; Crabtree, G. R. *Nature* **1996**, *382*, 822-826.
- (13) Arnone, M. I.; Davidson, E. H. *Development* **1997**, *124*, 1851-1864.
- (14) Jones, K. A.; Peterlin, B. M. *Ann. Rev. Biochem.* **1994**, *63*, 717-743.
- (15) Ferredamare, A. R.; Pognonec, P.; Roeder, R. G.; Burley, S. K. *EMBO J.* **1994**, *13*, 180-189.
- (16) Kodandapani, R.; Pio, F.; Ni, C. Z.; Piccialli, G.; Klemsz, M.; McKercher, S.; Maki, R. A.; Ely, K. R. *Nature* **1996**, *380*, 456-460.

- (17) Love, J. J.; Li, X. A.; Case, D. A.; Giese, K.; Grosschedl, R.; Wright, P. E. *Nature* **1995**, 376, 791-795.
- (18) Ghosh, G.; Vanduyne, G.; Ghosh, S.; Sigler, P. B. *Nature* **1995**, 373, 303-310.
- (19) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, 252, 809-817.
- (20) Geiger, J. H.; Hahn, S.; Lee, S.; Sigler, P. B. *Science* **1996**, 272, 830-836.
- (21) Nikolov, D. B.; Chen, H.; Halay, E. D.; Usheva, A. A.; Hisatake, K.; Lee, D. K.; Roeder, R. G.; Burley, S. K. *Nature* **1995**, 377, 119-128.
- (22) Sjøttem, E.; Andersen, C.; Johansen, T. *J. Mol. Biol.* **1997**, 267, 490-504.
- (23) Suzuki, M.; Brenner, S. E.; Gerstein, M.; Yagi, N. *Prot. Eng.* **1995**, 8, 319-328.
- (24) Pabo, C. O.; Sauer, R. T. *Ann. Rev. Biochem.* **1992**, 61, 1053-1095.
- (25) Burley, S. K. *Curr. Op. Struct. Biol.* **1994**, 4, 3-11.
- (26) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, 71, 1223-1237.
- (27) Kissinger, C. R.; Liu, B. S.; Martinblanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, 63, 579-590.
- (28) TuckerKellogg, L.; Rould, M. A.; Chambers, K. A.; Ades, S. E.; Sauer, R. T.; Pabo, C. O. *Structure* **1997**, 5, 1047-1054.
- (29) Ma, P. C. M.; Rould, M. A.; Weintraub, H.; Pabo, C. O. *Cell* **1994**, 77, 451-459.
- (30) Kim, Y. C.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, 365, 512-520.
- (31) Kim, J. L.; Nikolov, D. B.; Burley, S. K. *Nature* **1993**, 365, 520-527.
- (32) Park, C. M.; Campbell, J. L.; Goddard, W. A. *Proc. Natl. Acad. Sci. USA* **1992**, 89, 9094-9096.
- (33) Park, C.; Campbell, J. L.; Goddard, W. A. *Proc. Natl. Acad. Sci. USA* **1993**, 90, 4892-4896.
- (34) Pomerantz, J. L.; Pabo, C. O.; Sharp, P. A. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 9752-9756.
- (35) Pomerantz, J. L.; Sharp, P. A.; Pabo, C. O. *Science* **1995**, 267, 93-96.

- (36) Greisman, H. A.; Pabo, C. O. *Science* **1997**, 275, 657-661.
- (37) Choo, Y.; Sanchezgarcia, I.; Klug, A. *Nature* **1994**, 372, 642-645.
- (38) Isalan, M.; Klug, A.; Choo, Y. *Biochem.* **1998**, 37, 12026-12033.
- (39) Beal, P. A.; Dervan, P. B. *Science* **1991**, 251, 1360-1363.
- (40) Moser, H. E.; Dervan, P. B. *Science* **1987**, 238, 645-650.
- (41) Greenberg, W. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, 117, 5016-5022.
- (42) Best, G. C.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, 117, 1187-1193.
- (43) Strobel, S. A.; Doucetestamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, 254, 1639-1642.
- (44) Maher, L. J.; Dervan, P. B.; Wold, B. *Biochem.* **1992**, 31, 70-81.
- (45) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, 241, 456-459.
- (46) Mouscadet, J. F.; Carteau, S.; Goulaouic, H.; Subra, F.; Auclair, C. *J. Biol. Chem.* **1994**, 269, 21635-21638.
- (47) Bouziane, M.; Cherny, D. I.; Mouscadet, J. F.; Auclair, C. *J. Biol. Chem.* **1996**, 271, 10359-10364.
- (48) Helene, C. *Anti-Cancer Drug Design* **1991**, 6, 569-584.
- (49) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, 114, 1895-1897.
- (50) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, 254, 1497-1500.
- (51) Hanvey, J. C.; Pepper, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G.; Carter, S. G.; Bruckenstein, D. A.; Boyd, A. L.; Noble, S. A.; Babiss, L. E. *Science* **1992**, 258, 1481-1485.
- (52) Buchardt, O.; Egholm, M.; Berg, R. H.; Nielsen, P. E. *Trends Biotech.* **1993**, 11, 384-386.
- (53) Eriksson, M.; Nielsen, P. E. *Nat. Struct. Biol.* **1996**, 3, 410-413.

- (54) Chaires, J. B. *Biopoly* **1998**, *44*, 201-215.
- (55) Ikeda, R. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 296-297.
- (56) Rye, H. S.; Yue, S.; Wemmer, D. E.; Quesada, M. A.; Haugland, R. P.; Mathies, R. A.; Glazer, A. N. *Nucl. Acids Res.* **1992**, *20*, 2803-2812.
- (57) Glazer, A. N.; Rye, H. S. *Nature* **1992**, *359*, 859-861.
- (58) Chaires, J. B.; Leng, F. F.; Przewloka, T.; Fokt, I.; Ling, Y. H.; PerezSoler, R.; Priebe, W. *J. Med. Chem.* **1997**, *40*, 261-266.
- (59) Leng, F. F.; Priebe, W.; Chaires, J. B. *Biochem.* **1998**, *37*, 1743-1753.
- (60) Wemmer, D. E.; Dervan, P. B. *Curr. Op. Struct. Biol.* **1997**, *7*, 355-361.
- (61) Geierstanger, B. H.; Wemmer, D. E. *Ann. Rev. Biohys. Biomol. Struct.* **1995**, *24*, 463-469.
- (62) Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. *J. Mol. Biol.* **1997**, *271*, 244-257.
- (63) Johann, T. W.; Barton, J. K. *Phil. Trans. Roy. Soc. London Ser. a-Math. Phys. Eng. Sci.* **1996**, *354*, 299-324.
- (64) Hudson, B. P.; Barton, J. K. *J. Am. Chem. Soc.* **1998**, *120*, 6877-6888.
- (65) Eriksson, M.; Leijon, M.; Hiort, C.; Norden, B.; Graslund, A. *J. Am. Chem. Soc.* **1992**, *114*, 4933-4934.
- (66) Tuite, E.; Lincoln, P.; Norden, B. *J. Am. Chem. Soc.* **1997**, *119*, 239-240.
- (67) Bailly, C.; Chaires, J. B. *Bioconj. Chem.* **1998**, *9*, 513-538.
- (68) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385.
- (69) Bailly, C.; Henichart, J. P. *Bioconj. Chem.* **1991**, *2*, 379-393.
- (70) Pelton, J. G.; Wemmer, D. E. *J. Biol. Struct. Dyn.* **1990**, *8*, 81-97.
- (71) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575.
- (72) Wade, W. S.; Dervan, P. B. *Abs. Pap. Am. Chem. Soc.* **1988**, *196*, 235-ORGN.

- (73) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586-7590.
- (74) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochem.* **1993**, *32*, 11385-11389.
- (75) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.
- (76) White, S.; Baird, E. E.; Dervan, P. B. *Biochem.* **1996**, *35*, 12532-12537.
- (77) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471.
- (78) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.
- (79) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600-1607.
- (80) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534-3535.
- (81) Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104-109.
- (82) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, 111-115.
- (83) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899.
- (84) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.
- (85) Cho, J. Y.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10389-10392.
- (86) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1996**, *3*, 369-377.
- (87) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564-7566.
- (88) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
- (89) Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Ang. Chem. Int. Ed. Eng.* **1996**, *35*, 1487-1489.

- (90) Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6778-6779.

Chapter 2:

Sequence-Specific Detection of Double-Stranded DNA

2.1 Abstract

Sequence-specific fluorescent DNA ligands could be important reagents for applications in diagnostics and molecular biology. Such reagents have potentially distinct advantages over conventional hybridization based approaches. This chapter presents preliminary investigations into the use of polyamides as sequence-specific DNA sensitive fluorophores. Three distinct approaches were undertaken. First, polyamide cyanine dye conjugates were developed that fluoresce upon binding to double stranded DNA. Such compounds bind DNA with nanomolar affinities and produce large fluorescence enhancements upon binding. Second, preliminary investigations were undertaken to use energy transfer system between fluorescent dyes to detect binding of polyamides to adjacent DNA binding sites. Lastly, a system designed to form pyrene excimers through binding of polyamides to adjacent sites was explored.

2.2 Detection with Cyanine Dye Conjugates

2.2.1 Background

Most fluorescent assays require denaturation of the target DNA sequence before detection.^{1,2} As pyrrole-imidazole polyamides bind to double stranded DNA, this step

would be unnecessary. Secondly, the recognition of native double stranded DNA with fluorescent minor groove binding polyamides could ultimately avoid the amplification steps necessary for most DNA detection methods. Coupling the recognition of the polyamide to the specific generation of a fluorescent signal upon binding to a target DNA site would result in assays with reduced background compared to conventional assays. Such an approach requires the design and synthesis of appropriate polyamide fluorophore conjugates that retain high specificity for targeted DNA match sites, and result in high levels of fluorescence upon binding to the target site.

With these criteria in mind, we have designed a series of conjugates of pyrrole imidazole polyamides with fluorescent cyanine dyes.³ These dyes are quite unique in their properties of extreme fluorescent enhancement upon binding to double stranded DNA.⁴ Fluorescent enhancements can range from 200 to 3,000 fold, while the classic intercalative fluorophore, ethidium bromide, shows an enhancement of only approximately 10 to 40 fold.⁵ This enhancement is most likely the result of the conformational restriction of the dye in the bound state, allowing resonance of the positive charge along the length of the dye molecule and preventing radiational energy loss.⁶ This restriction can result from full intercalation of the dye molecule or through groove binding modes. The structure of the most common of these cyanine dyes, thiazole orange, is shown in figure 1.

Another promising feature of the cyanine dyes is that the linker length and electronic character of the dye substituents can be used to tune the dye's absorption and emission from orange to violet.^{3,5} Using multiple dye molecules, this class of dyes has been used in the detection of single molecules of DNA.^{7,8} Such single molecule detection avoids potential contamination from amplification procedures and is also suitable for direct physical optical mapping applications.⁹⁻¹¹ With sensitive detection techniques, it is now possible to detect single fluorophore molecules.¹² Castro and Williams, for instance, have applied this technology to detect specific DNA sequences in unamplified genomic DNA

through hybridization of fluorescently labeled PNA probes using an ultrasensitive laser fluorescence system.¹³

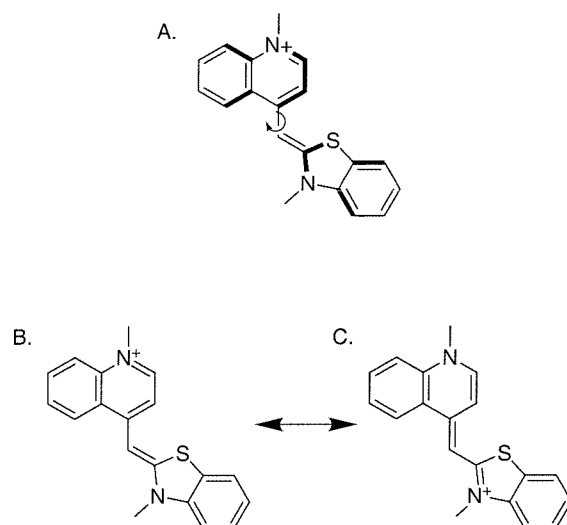


Figure 1. Conformation and resonance of thiazole orange. A. Representation of the solution conformation of thiazole orange, with the aromatic rings out of plane and free to rotate lead to radiative loss of energy. B. and C. Planarization of the rings through interaction with DNA allows resonance of charge across the compound, leading to enhanced fluorescence and prevents radiative energy loss.

Three cyanine dye molecules were synthesized with butyric acid moieties for conjugation to the polyamides. These dyes and their synthesis are shown in figure 3. They are thiazole orange (TO), thiazole yellow (TY), and thiazole blue (TAB). Literature reports of these dyes shows that dyes with these core structures undergo significant fluorescence enhancement upon binding DNA.¹⁴⁻¹⁶ Values for the excitation wavelength, emission wavelength, and fluorescence enhancement upon binding to DNA are shown below. Reported literature values for fluorescence enhancements vary greatly, mostly as a result of the fact that they are a ratio of a large number to a small number; the dyes are virtually non-fluorescent in the absence of DNA.

Table 1. Fluorescence properties of three cyanine dyes compared to EthBr.¹⁴⁻¹⁶

Dye	Absorbance (ϵ (M ⁻¹ cm ⁻¹))	Emission (DNA bound)	Fluorescence Enhancement
Ethidium Bromide	516 (3,700)	616	13-40
Thiazole Orange	512 (46,900)	527	2,600-3,000
Thiazole Yellow	462 (58,100)	481	892
Thiazole Blue	639 (61,200)	658	388

These dyes were conjugated to a well defined polyamide, ImPyPyPy- γ -PyPyPyPy- β -Dp,¹⁷ through linkage at either the C-terminus (for thiazole orange only) or at the γ -aminobutyric acid turn residue (all three dyes). The dye conjugates were characterized for their ability to bind DNA sequence specifically by two techniques. First, spectroscopic investigations followed fluorescence enhancements upon binding to various DNA oligonucleotides and oligoduplexes. Second, quantitative DNase I footprinting was used to analyze each compound's affinity and specificity for DNA.

2.2.2 Design

The NMR structure of the thiazole orange dimer (TOTO) was examined to determine the starting design of the polyamide conjugates. The linking region of TOTO is clearly in the minor groove of the DNA, making it an attractive point of attachment to a minor groove binding polyamide.⁶ Using molecular modeling, a simple butyric acid moiety was chosen as a linker to the polyamide. Placement of the dye at either the turn or C-terminus of the polyamide allows for flexibility in the design of the polyamide conjugates. In addition, future studies can optimize the linkage between the two halves of the bifunctional molecule. Dimeric cyanine dyes, for instance, are remarkably sensitive to single methylene differences in linkers connecting the dyes.¹⁸ While the monomeric cyanine dyes have little sequence preference, the dimeric cyanine dye thiazole orange was

shown to prefer 5'-CTAG-3' binding sites.¹⁹ Any sequence preference may also effect the polyamide conjugates.

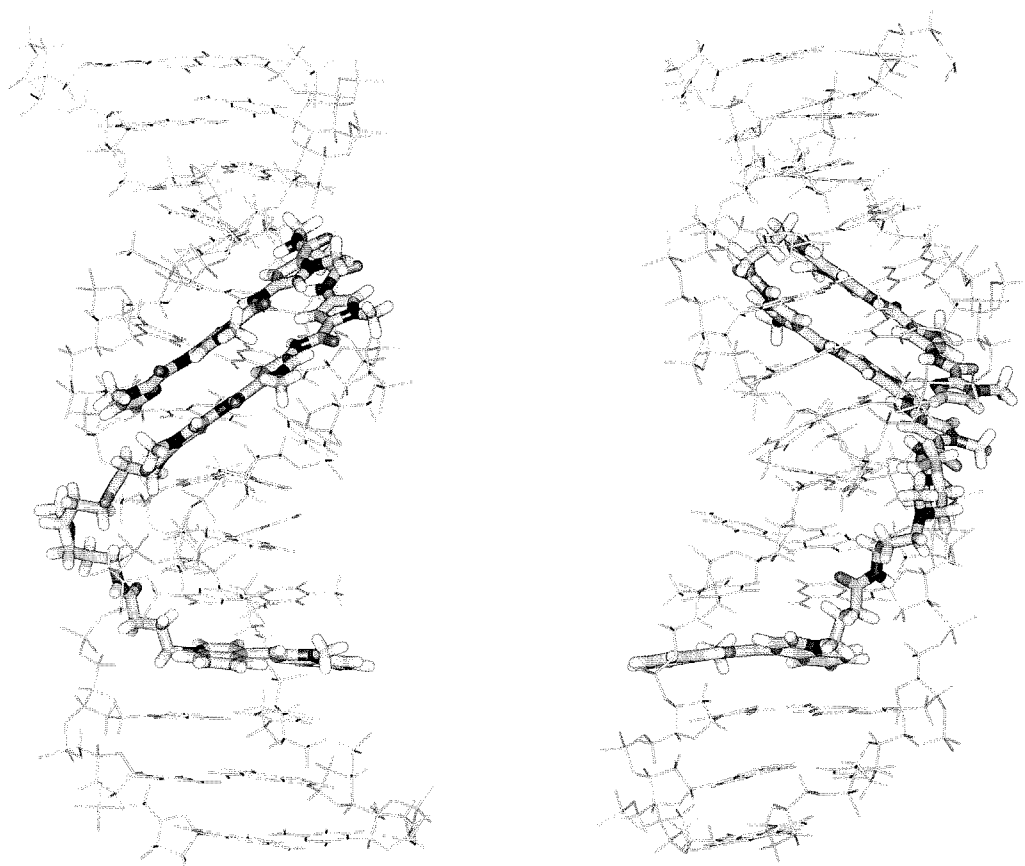


Figure 2. Model of ImPyPy- γ -PyPyPy- β -Dp-TO. Generated using InsightII from coordinates of ImPyPy- γ -PyPyPy-Dp²⁰ and thiazole orange dimer (TOTO).⁶ The linker region was minimized using the CVFF forcefield, while the polyamide and intercalator subunits were held fixed.

2.2.3 *Synthesis of Compounds*

After initial unsuccessful attempts to alkylate the dimethylaminopropylamine tail of a polyamide with thiazole orange propyl iodide,¹⁵ synthetic efforts towards the preparation of cyanine dye conjugates of polyamides focused on amide bond formation between cyanine dye acids and polyamide amines. Towards this end, the three cyanine dyes shown

in figure 3 were synthesized bearing butyric acid side chains. Neat alkylation of either lepidine or 4-picoline with ethyl-4-bromobutyrate gave pyridinium salts in excellent yield. The dye esters were synthesized by well precededented condensation reactions,³ then saponified to give the appropriate dye acids. Importantly, it was found that precipitation of the dye acid often led to purification of insoluble salts of the dye rather than the free acid. Purification of the compounds by reversed phase silica gel chromatography eluting with 0.1% trifluoroacetic acid 25% acetonitrile afforded a more soluble form of the dye, presumably the free acid, trifluoroacetate salt.

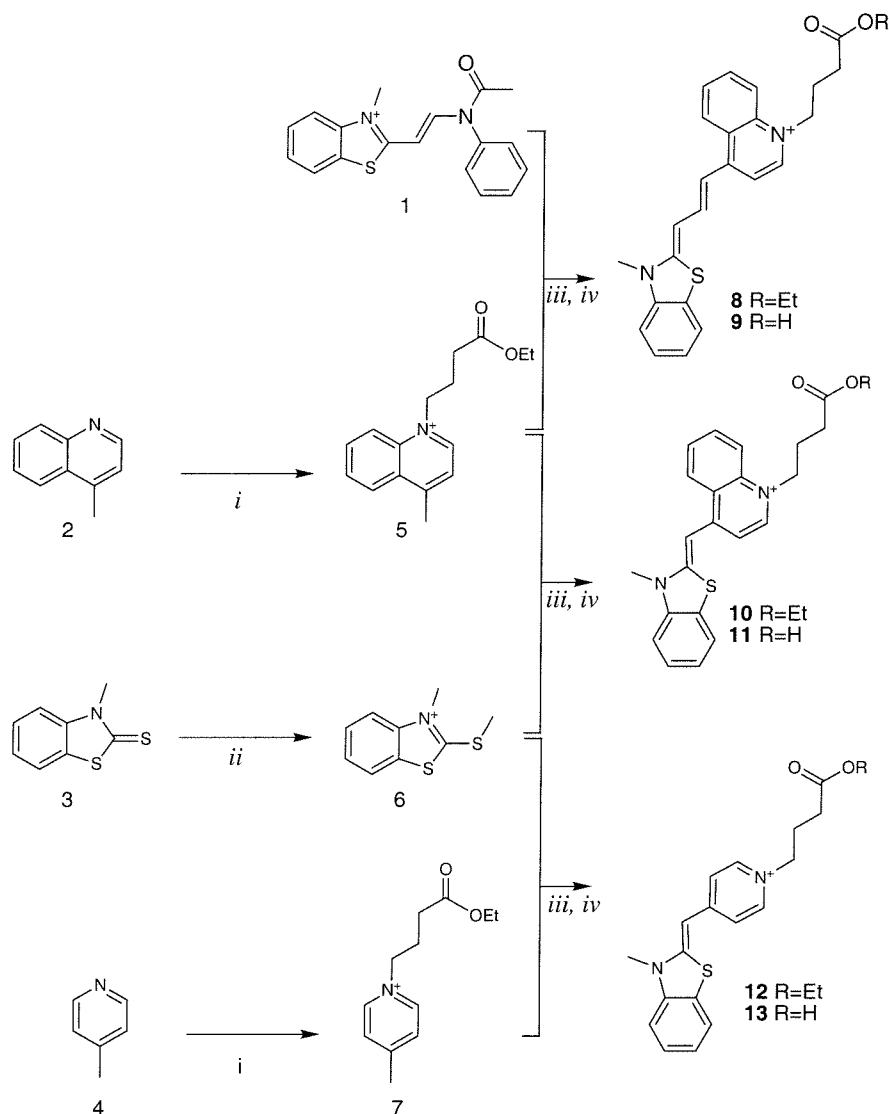


Figure 3. Synthesis of cyanine dye derivatives 9, thiazole blue (TB); 10, thiazole orange (TO); 13, thiazole yellow (TY). *i*) ethyl-4-bromobutyrate, 120 °C, 95-99%; *ii*) MeI, EtOH, ;*iii*) EtOH, Et₃N reflux, 60-95%; *iv*) KOH, MeOH, H₂O 75-95%.

Efforts to conjugate the dyes to polyamides did not proceed smoothly. Several difficulties were encountered in the synthesis. The poor solubility of the dye acids was overcome by preparing the free acid. Some coupling reagents, however, caused precipitation of the dye. Second, the retention times of conjugates were often similar to that of the free dye or the activated form of the dye. This precluded the use of large excesses of dye in coupling reactions. Also, the solubility of conjugates depended strongly on the net

charge of the conjugate. Compounds with only a single positive charge had poor solubility characteristics. Lastly, the HPLC recoveries of dye conjugates were notably low, even when using a C4 reversed phase rather than a C18 reversed phase column. Ultimately, use of the free acid and activation with HBTU gave fairly reproducible results in coupling reactions. The solubility properties of the prepared conjugates were also somewhat problematic, particularly for the thiazole orange conjugates. Purified aliquots of the thiazole orange conjugates were not easily resuspended in water, but were readily resuspended in DMSO. All compounds could be obtained as 50 μ M solutions in 10% DMSO/H₂O. The thiazole yellow conjugate showed much better solubility properties, with aqueous solutions readily obtained at 50 μ M.

Counterintuitively, synthesis of the conjugates prepared at the more hindered α amine of the R α,γ -diaminobutyric acid residue proceeded more smoothly than with the C-terminal propylamine compounds. This could possibly be a result of an impurity in the prepared polyamide amines or, perhaps, a conformational problem. Efforts to synthesize conjugates with polyamides cleaved from the resin with shorter alkyl diamines (ethylene diamine, 1,3-propane diamine, and 1,4-butane diamine) gave poor yields under identical conditions.

Polyamide conjugates were purified to give single peaks by analytical HPLC, and the structures confirmed by mass spectrometry (all compounds) and by ¹H-NMR (compound **14**). Electrospray ionization gave better results for these compounds than MALDI, presumably by prevention of photooxidation. Products with mass 16 higher than calculated were often seen for these conjugates by MALDI-TOF, but were present in significantly smaller amounts using electrospray ionization.

A second route was also briefly explored toward the synthesis of these conjugates. We sought to conjugate the quinidinium half dye to the polyamide and form the dye after polyamide purification. It was hoped that this would allow for easier HPLC separations, as the dye would only be present on the conjugate. Saponification of **5** with potassium

hydroxide in methanol/water led to formation of a dark red, unidentified product. The ester could be cleaved by acidic hydrolysis in refluxing concentrated HCl, but all attempts at activating or coupling this acid led to the formation of the deep red color, presumably formation of a symmetric cyanine dye.

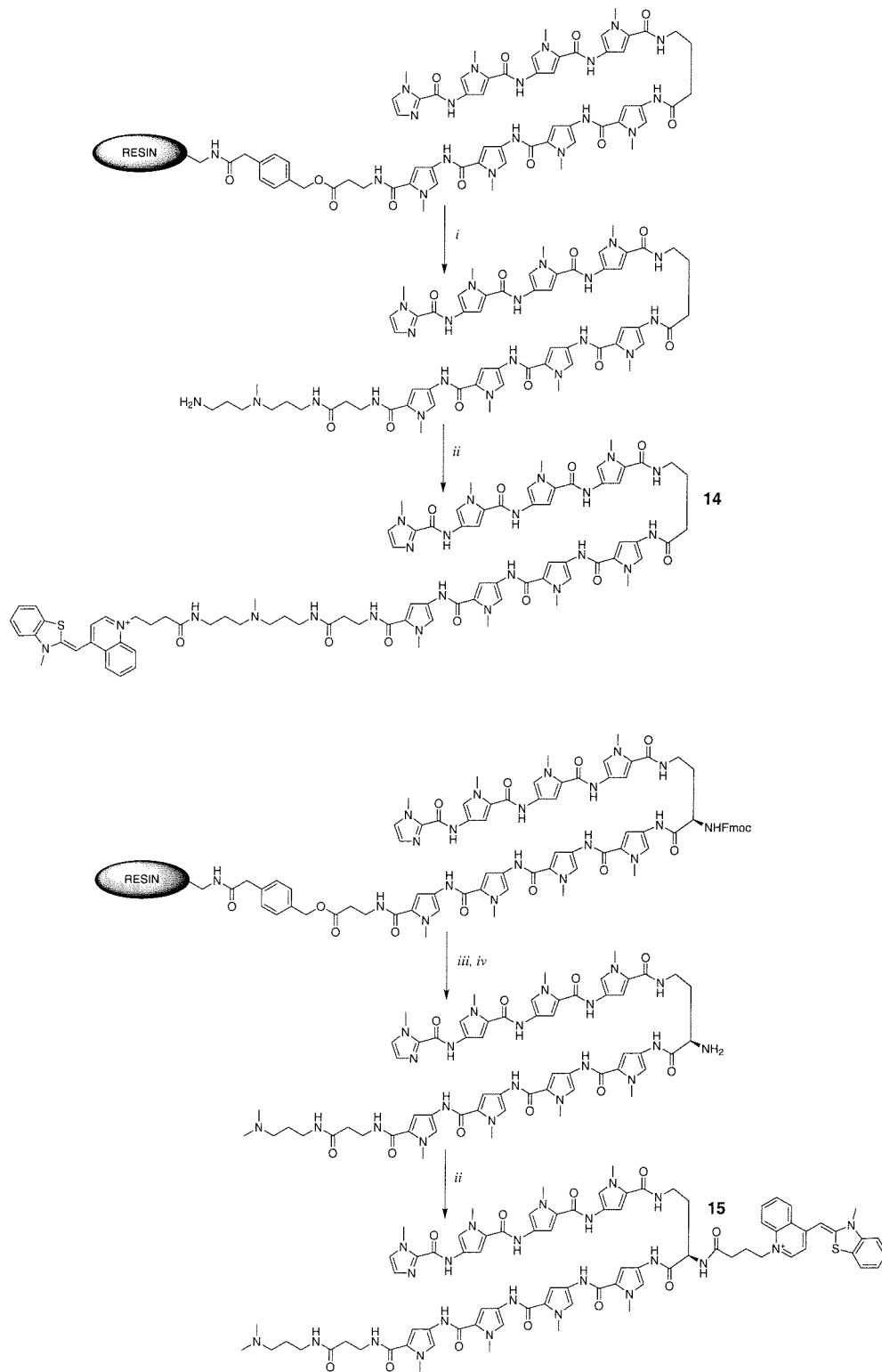


Figure 4. Synthesis of cyanine dye conjugates. *i*) *N*-methyl-*N,N*-dipropylamine, 55 °C, HPLC purification. *ii*) **11**, HBTU, DIEA, DMF; *iii*) 20% piperidine; *iv*) *N,N*-Dimethylaminopropylamine.

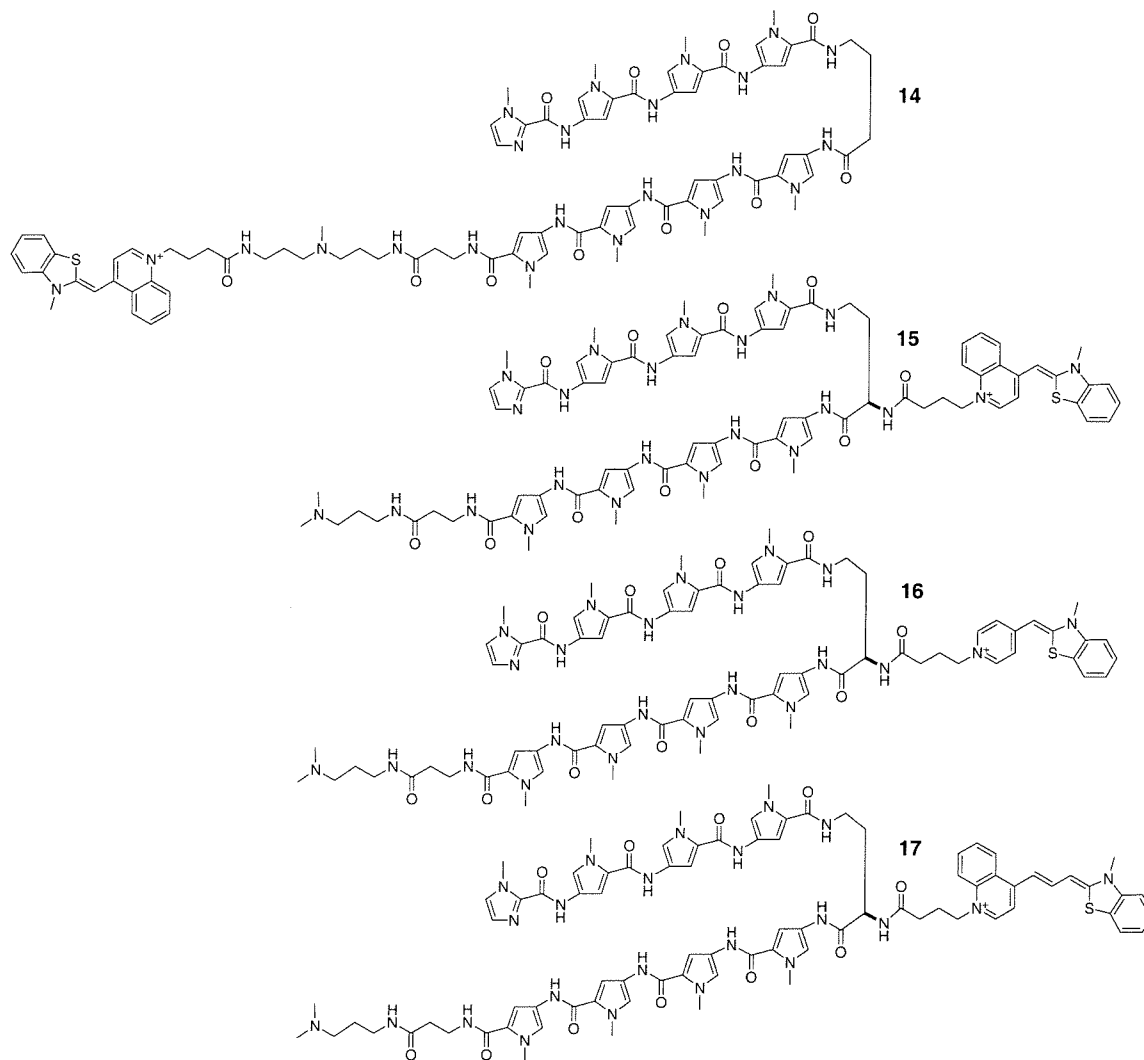


Figure 5. Structures of the polyamide cyanine dye conjugates used in this study.

2.2.4 Results and Discussion

The absorbance spectra of these cyanine conjugates were as expected with major absorbance at approximately 310nm corresponding to the polyamide absorbance, and a large absorbance band at the expected wavelength for each dye. The absorbance spectra of the thiazole orange conjugates are shown in figure 6 while the thiazole blue and yellow conjugates are shown in figure 7.

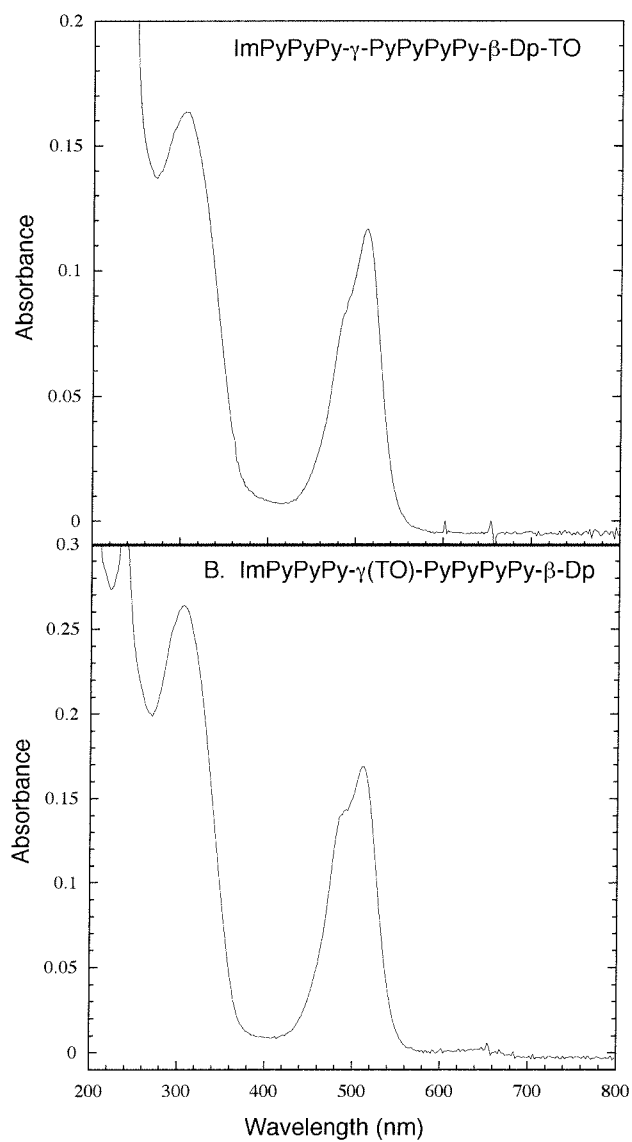


Figure 6. Absorbance spectra of thiazole orange conjugates. 14, upper, and 15, lower. Both spectra were recorded in water in a 1 cm quartz cuvette with 1.7 and 3.6 μ M solutions respectively.

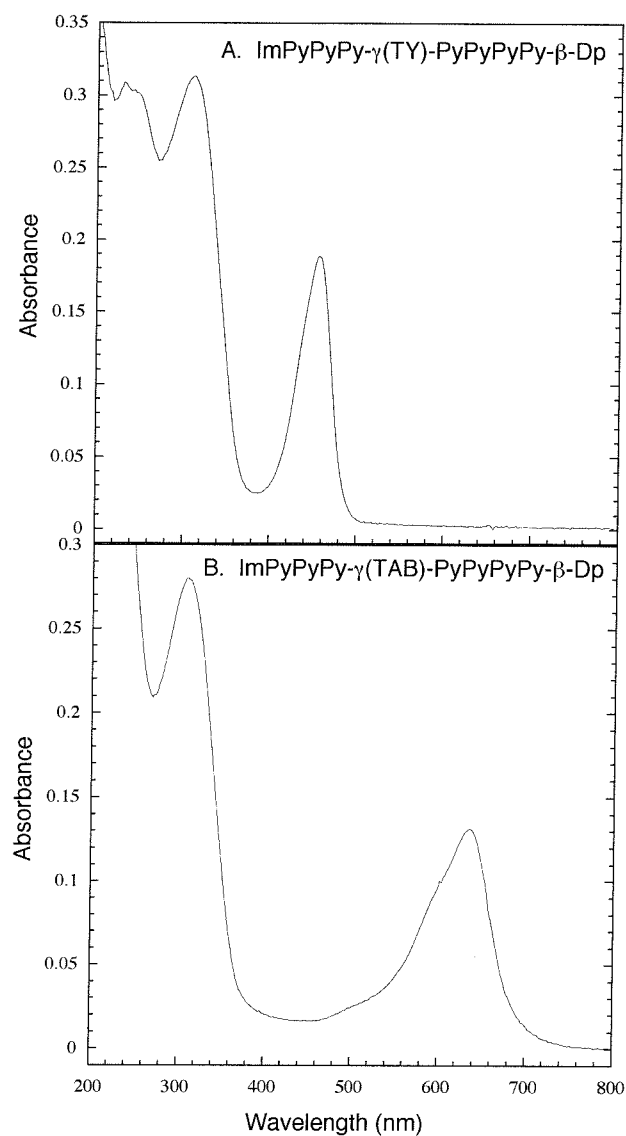
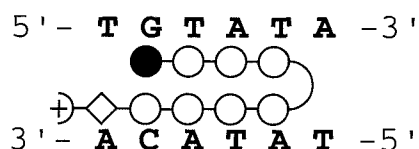


Figure 7. Absorbance spectra of thiazole yellow and thiazole blue derivatives. A. Thiazole yellow conjugate, **16**. B. Thiazole blue conjugate, **17**. Both spectra were recorded in water in a 1 cm quartz cuvette with 4.5 and 4.1 μM solutions respectively.

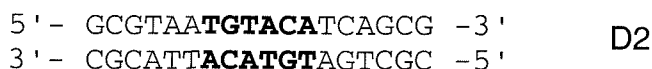
The conjugates are essentially non-fluorescent, but the conjugates fluoresce upon binding to double stranded DNA. Steady-state fluorescence spectra were taken of the titration of conjugates with oligoduplexes containing polyamide match binding sites and various mismatches. The oligonucleotides used in these experiments are shown in figure 8.



Match Site



Single Base Mismatch



Mismatch Oligo



Bulge Oligo



Single Strand



Figure 8. Oligonucleotides and duplexes used for fluorescence experiments.

These conjugates bind DNA with little sequence specificity at the concentrations necessary for these spectroscopic studies. Figure 9 shows the fluorescence of ImPyPyPy- γ -PyPyPyPy- β -Dp-TO with the various oligoduplexes. This was the best specificity that we were able to achieve. The ability to detect the bound species significantly limits the observable specificity of such compounds. In comparison to the DNase I footprinting titrations, there are several key differences. First, in the fluorescent experiment, only the bound species is detected. The DNA target must, therefore, be present in concentrations equal to or greater than the detection limit of the fluorometer used in these studies.

Presence of the DNA target at nanomolar or higher concentrations also removes the working assumption from the footprinting calculations that the polyamide concentration remains unchanged through a titration.

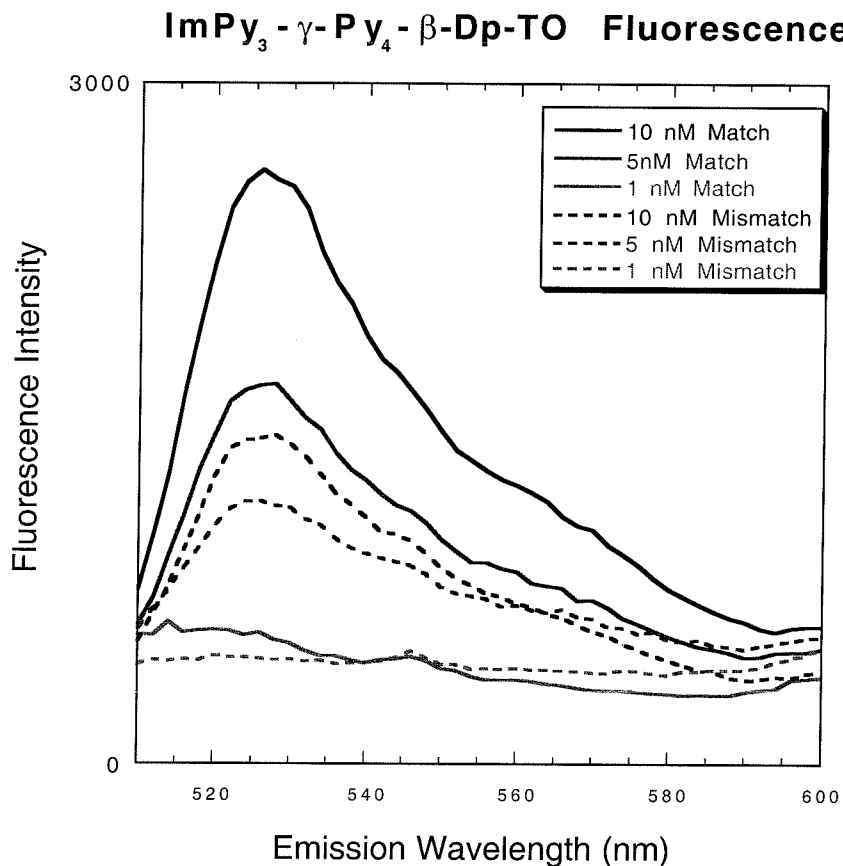


Figure 9. DNA detection by compound 14, ImPyPyPy-γ-PyPyPyPy-β-Dp-TO in the presence of match (D1) or mismatch (D2) oligoduplexes. Polyamide concentrations are as shown. Oligoduplexes were present at 100 nM. Solution conditions were 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, and pH 7.0. Samples were excited at 490 nm. Data were collected at 2 nm steps with 10 second integration in a 1 cm cuvette. Emission and excitation slit widths were set to 5 nm.

Fluorescence studies also suggest that these compounds form sequence-specific complexes with DNA. Titration of 100 nM oligoduplex D1 with ImPyPyPyPy- γ -PyPyPyPyPy- β -Dp-TO shows saturation of binding after formation of a 1:1 complex.

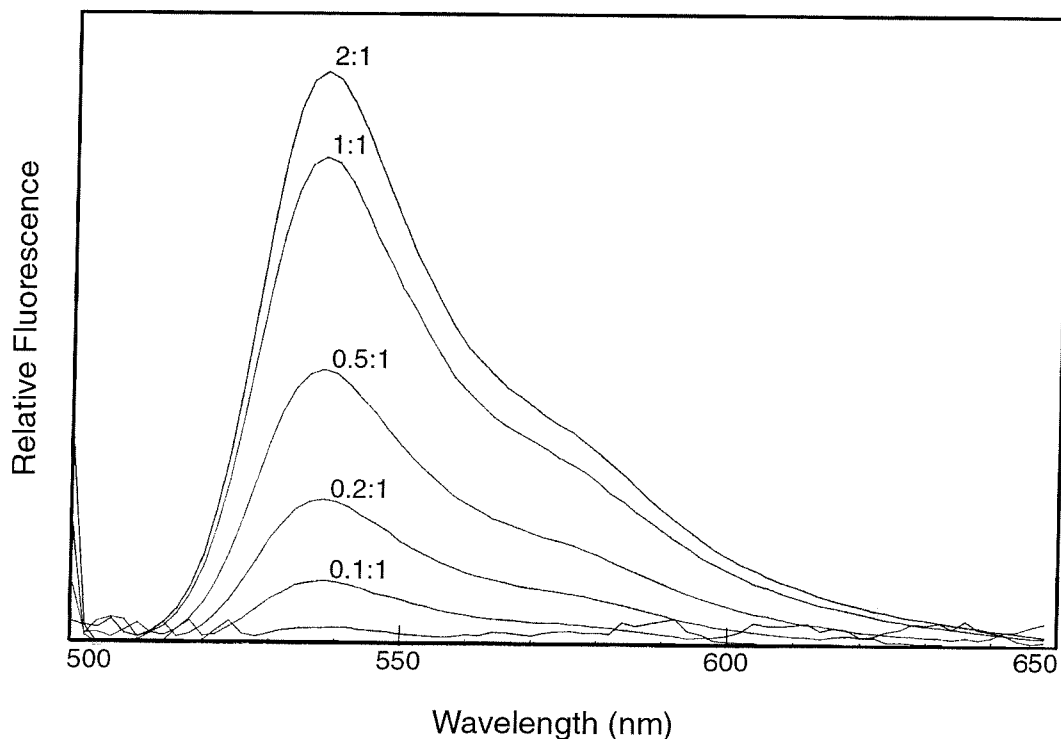


Figure 10. Titration of ImPyPyPyPy- γ -PyPyPyPyPy- β -Dp-TO with Oligoduplex D1. Oligoduplex concentration is 100 nM in all experiments. Polyamide **14** is titrated at concentrations of 5, 10, 20, 50, 100, and 200 nM. Spectra were acquired with excitation at 490 nm in a 1cm cuvette. Emission spectra was collected at 2 nm increments with 10 sec integration at each point.

Polyamide cyanine dye conjugates discriminate between single and double stranded DNA. As shown in figures 11-14, the cyanine dye conjugates excel in the discrimination between single stranded oligo S1, and the corresponding duplexes D1 through D4. Interestingly, the compounds show distinct differences in their interactions with duplex DNA. Linkage of the thiazole orange to the γ -aminobutyric acid residue rather than to the C-terminus results in a roughly eight-fold reduction in fluorescence intensity at the same dye concentration. The thiazole blue conjugate, **17**, shows a similarly reduced

fluorescence intensity. Interestingly, the thiazole yellow conjugate prepared at the turn residue shows a fluorescence enhancement similar to the C-terminal thiazole orange conjugate, **14**. Steric interactions could preclude the bulky thiazole blue and orange dyes from making positive contacts with the DNA when placed at the turn position. The similarity of the thiazole yellow and orange conjugates was also unexpected as the thiazole yellow was reported to have a lower fluorescence enhancement.

The fluorescence spectra show some sequence dependence. Unlike the low concentration experiments, the thiazole orange conjugates show little difference in fluorescence intensity at the concentrations used in these experiments (250 nM conjugate, 500 nM DNA). The thiazole yellow conjugate, however, shows reproducible fluorescence differences with the various duplexes. There is not a clear explanation for this. It is possible that the complexes with the mismatch oligoduplexes allow for better orientation and binding of the dye. The thiazole blue conjugate, although showing only weak fluorescence, shows some marked sequence-dependent fluorescence changes. The complex formed with a bulged oligoduplex shows an approximately 15 nm shift in fluorescence emission. This suggests, perhaps that this dye preferentially intercalates into the bulge site. Understanding the details of such interactions will require structural analysis of the polyamide dye conjugates in the presence of various duplexes. NMR studies in particular should differentiate between intercalative and groove binding modes.

Binding of the conjugates occurs at concentrations lower than for the free dyes. Suggesting that the observed complexes are dependent on the polyamide interaction with the DNA as well as that of the intercalator moiety. In experiments not shown, the dye butyric acids were coupled to *N,N*-dimethylaminopropane. These monomeric dyes showed no fluorescence with DNA below micromolar concentrations.

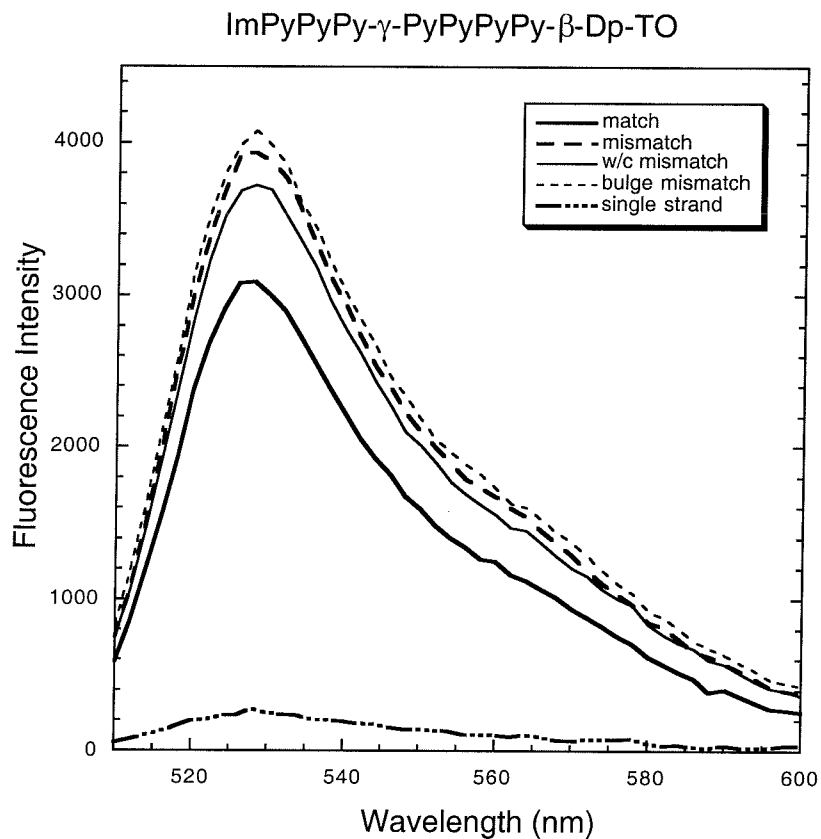


Figure 11. Fluorescence of ImPyPyPy- γ -PyPyPyPy- β -Dp-TO. Spectra are shown in the presence of oligoduplexes D1 (match), D2 (mismatch), D3 (w/c mismatch), D4 (bulge mismatch), and S1 (single strand). Polyamide conjugate was present in 250 nM concentration and the corresponding nucleic acid at 500 nM. Spectra were collected in a 0.5 cm cuvette with $\lambda_{\text{ex}} = 490$ nm. Data were collected from 510 to 600 nm in 2 nm increments with 10 sec integration at each wavelength. Excitation and emission slits were 5 nm. All spectra are corrected by subtraction of the spectrum of the free dye in the absence of DNA. No fluorescence was observed for the free conjugate.

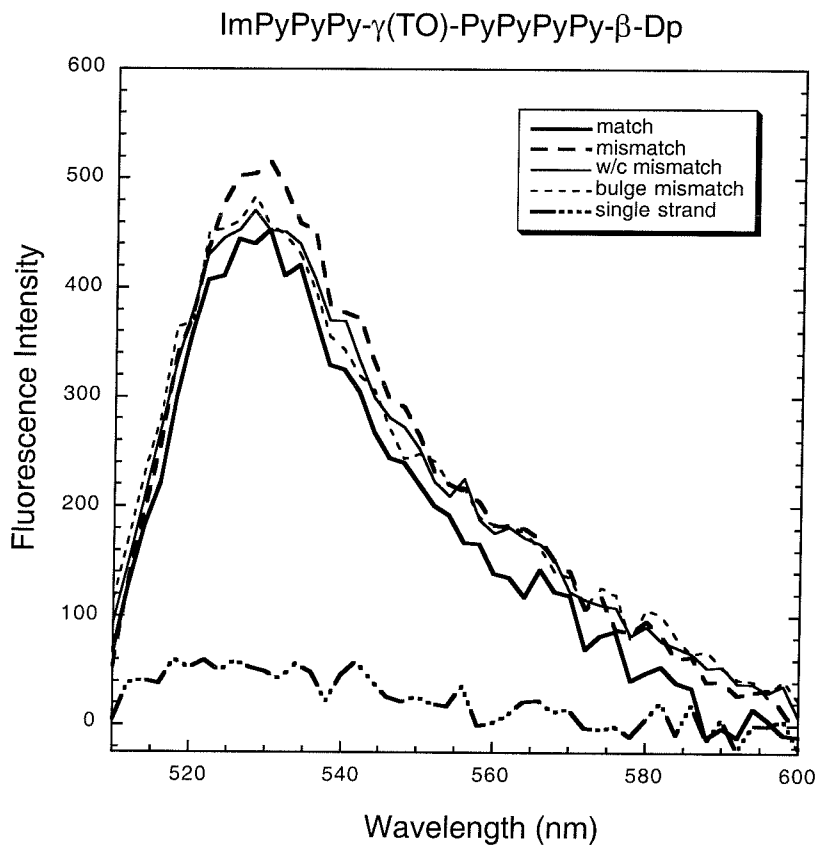


Figure 12. Fluorescence of ImPyPyPy- γ (TO)-PyPyPyPy- β -Dp (15) in the presence of oligoduplexes D1 (match), D2 (mismatch), D3 (w/c mismatch), D4 (bulge mismatch), and S1 (single strand). Polyamide conjugate was present in 250 nM concentration and the corresponding nucleic acid at 500 nM. Spectra were collected in a 0.5 cm cuvette with $\lambda_{\text{ex}} = 490$ nm. Data were collected from 510 to 600 nm in 2 nm increments with 10 sec integration at each wavelength. Excitation and emission slits were 5 nm. All spectra are corrected by subtraction of the spectrum of the free dye in the absence of DNA. No fluorescence was observed for the free conjugate.

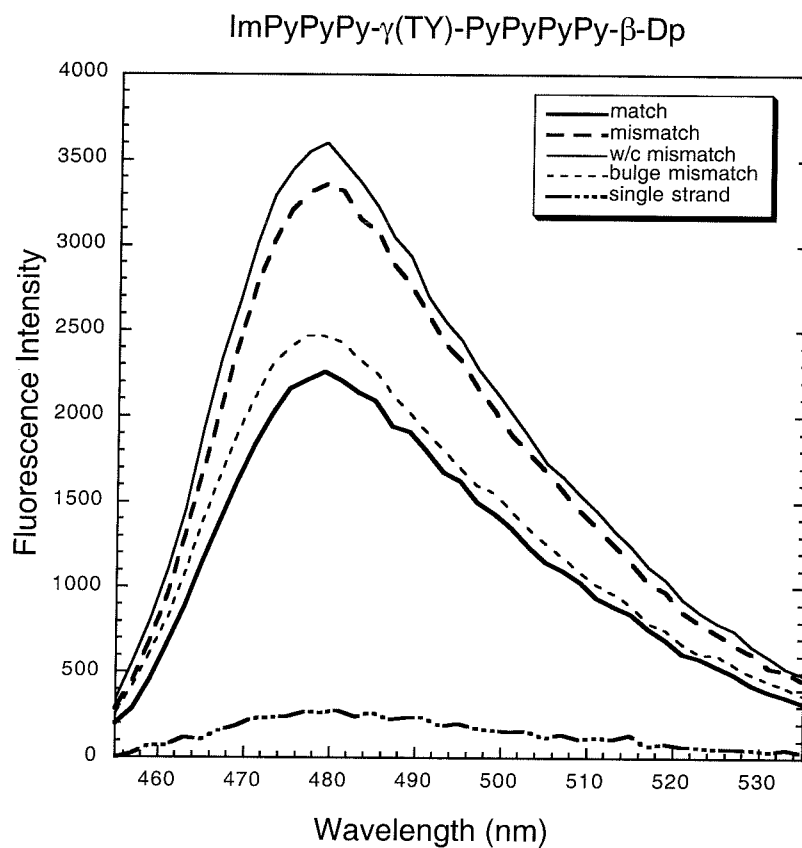


Figure 13. Fluorescence of ImPyPyPy- γ (TY)-PyPyPyPy- β -Dp (16) in the presence of oligoduplexes D1 (match), D2 (mismatch), D3 (w/c mismatch), D4 (bulge mismatch), and S1 (single strand). Polyamide conjugate was present in 250 nM concentration and the corresponding nucleic acid at 500 nM. Spectra were collected in a 0.5 cm cuvette with λ_{ex} = 440 nm. Data were collected from 455 to 535 nm in 2 nm increments with 10 sec integration at each wavelength. Excitation and emission slits were 5 nm. All spectra are corrected by subtraction of the spectrum of the free dye in the absence of DNA. No fluorescence was observed for the free conjugate.

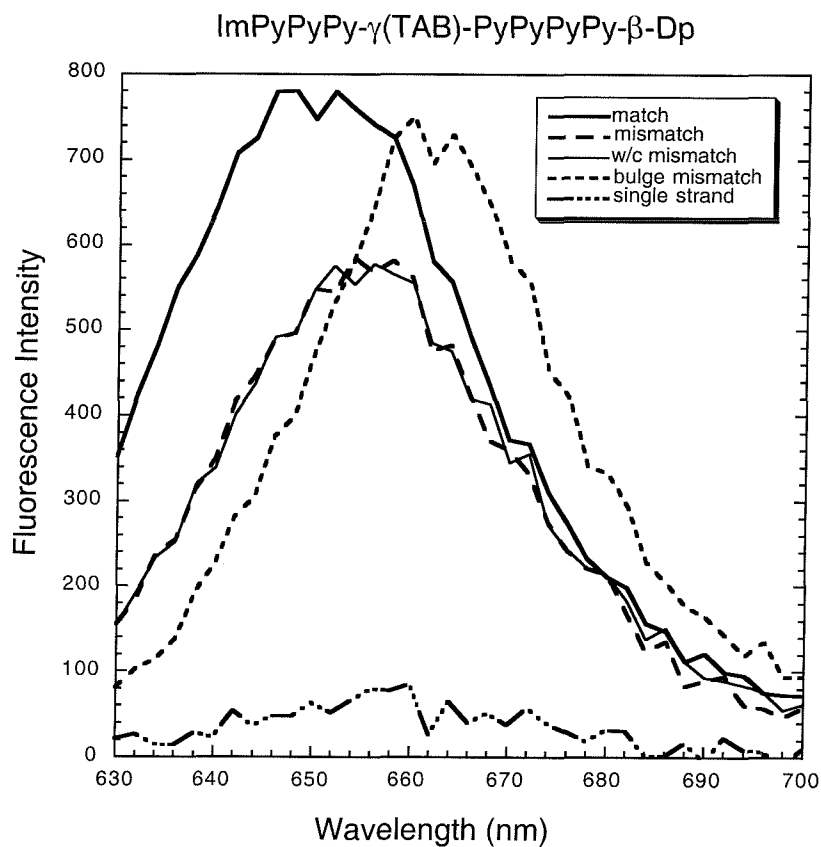


Figure 14. Fluorescence of ImPyPyPy- γ (TAB)-PyPyPyPy- β -Dp (17) in the presence of oligoduplexes D1 (match), D2 (mismatch), D3 (w/c mismatch), D4 (bulge mismatch), and S1 (single strand). Polyamide conjugate was present in 250 nM concentration and the corresponding nucleic acid at 500 nM. Spectra were collected in a 0.5 cm cuvette with $\lambda_{\text{ex}} = 620$ nm. Data were collected from 630 to 700 nm in 2 nm increments with 10 sec integration at each wavelength. Excitation and emission slits were 5 nm. All spectra are corrected by subtraction of the spectrum of the free dye in the absence of DNA. No fluorescence was observed for the free conjugate.

The apparent equilibrium association constants of these compounds were determined by quantitative DNase I footprinting. Compounds were footprinted on a

restriction fragment on which the parent hairpin polyamide had been footprinted, pJT8.¹⁷ Remarkably, addition of the dye to the polyamide results in little change in binding affinity for the conjugate. However, the overall specificities of the conjugates were greatly reduced. Apparent equilibrium association constants are summarized in table 2. While the thiazole orange derivative **14** shows high affinity for its targeted match site, the single base mismatch, 5'AGTACT-3' is bound with nearly the same affinity.

Conjugation of a cyanine dye to the gaba turn reduces both the affinity and specificity of the resulting conjugate. The lowest concentrations studied showed similar results to the parent compounds, but the conjugates showed complete lack of specificity at much lower concentrations than was observed for the parent hairpin. It is possible that the relatively tight binding of the intercalator portion of the molecule “flattens” the energy differences achieved by the polyamides. One approach to overcoming this difficulty would be to conjugate these dyes to polyamides that bind with significantly higher affinities than the dye moiety. A second approach that was not fully explored would be to fully explore the linkage between the polyamide and the dye moiety.

Another issue in the design of future conjugates incorporating cyanine dyes is the sequence specificity of the dyes themselves. Literature reports for the monomeric cyanine dyes suggest that they are relatively neutral with respect to sequence preferences. However, for the TOTO dimer, distinct sequence specificities have been observed.¹⁹

Table 2. Apparent equilibrium association constants (M^{-1}) of cyanine dye conjugates.^a

Polyamide Conjugate	5'-ttAGTATTTtg-3'	5'-ttAGTACTtg-3'
ImPyPyPy- γ -PyPyPyPy- β -Dp ¹⁷	3.5×10^9	5.0×10^8
ImPyPyPy- γ -PyPyPyPy- β -Dp-TO (14)	2.4×10^9	1.0×10^9
ImPyPyPy- γ (TO)-PyPyPyPy- β -Dp (15)	2.9×10^7	$< 5 \times 10^6$
ImPyPyPy- γ (TY)-PyPyPyPy- β -Dp (16)	1.4×10^9	6.8×10^8
ImPyPyPy- γ (TAB)-PyPyPyPy- β -Dp (17)	5.6×10^8	3.2×10^8

^a Quantitative DNase I footprinting reactions were carried out on a 3'-labeled EcoRI/PvuII restriction fragment derived from plasmid pJT8 as described.¹⁷ Data represent single trials.

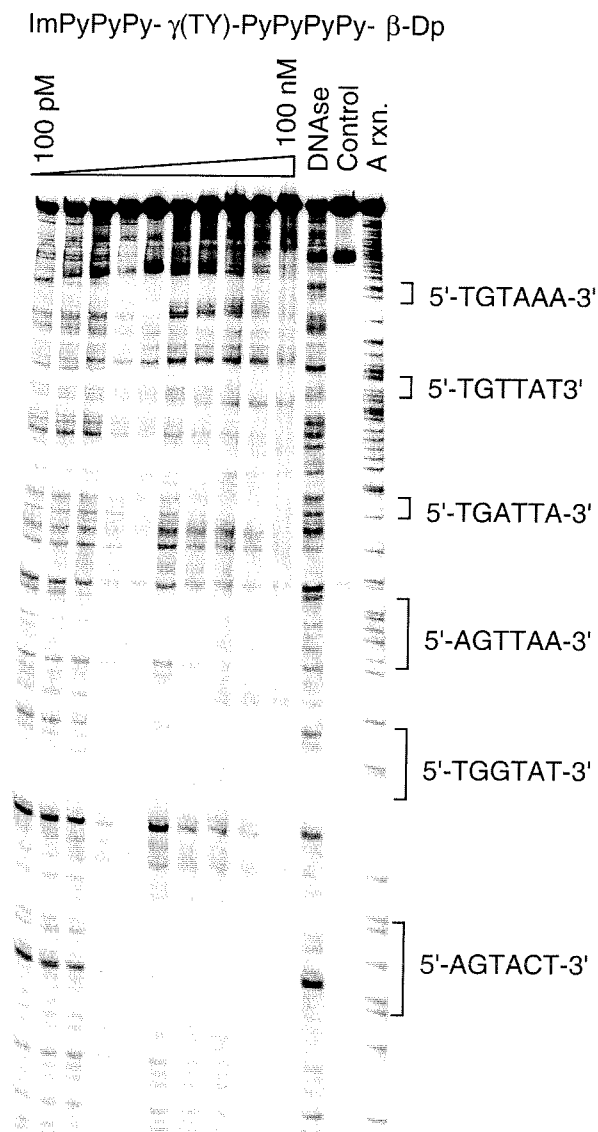


Figure 15. DNase I footprint titration of ImPyPyPy- γ (TY)-PyPyPyPy- β -Dp. Quantitative footprint titration on *EcoRI/PvuII* restriction fragment from pJT8. Reactions were performed at pH 7.0 in 400 μ L volumes with final concentrations of 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂. Polyamide concentrations used for titration were 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM. Lane marked DNase contains no polyamide. Control lane was not treated with DNase I. A rxn. denotes chemical sequencing lane.

2.2.5 Cyanine Conjugate Conclusions

Cyanine dyes can be conjugated to minor groove binding polyamides to give molecules capable of DNA dependent fluorescence at low nanomolar concentrations. These conjugates display large fluorescence enhancements upon binding to double stranded DNA, but not to single-stranded oligonucleotides. In a solution assay format with a conventional fluorometer, the specificity of these compounds is not sufficient to discriminate between match and mismatch containing oligoduplexes, nor to discriminate between duplexes with Watson-Crick mismatches or bulges. DNase I footprinting confirms that these conjugates have reduced specificities as compared to the parent hairpin polyamides. In other assay formats it may be possible to use second generation conjugates for sequence specific detection of DNA, particularly in a system with a more sensitive detector and a laser light source. Using polyamides with longer recognition sites should also improve the utility of such conjugates for sequence specific detection in a homogenous solution.²¹ Introduction of hydroxypyrrole residues should also improve the specificity of the conjugates.²²

2.3 Towards Detection Through Energy Transfer

2.3.1 Background

A number of approaches to the recognition of nucleic acids, or other biochemical entities have relied heavily on the phenomenon of Förster resonance energy transfer (FRET).²³ FRET is a process of non-radiative transfer of energy between two dyes with significant spectral overlap. The efficiency of transfer from donor to acceptor varies with a number of factors arising from an induced-dipole to induced-dipole interaction between the donor and acceptor. The efficiency of energy transfer, E , is dependent on the inverse 6th power of the distance R separating the dyes. The full equation to describe the phenomenon is $E = 1/(1+[R/R_0]^6)$, in which R_0 is the distance at which 50% energy transfer occurs. R_0 is dependent on a number of factors including spectral overlap $J(\lambda)$ and orientation κ^2 of the molecules: $R_0 = [8.79 \times 10^{-5} J(\lambda) \phi_D n^{-4} \kappa^2]^{1/6} \text{ \AA}$. ϕ_D is the quantum yield of the donor and n is the refractive index of the medium. Practically, in biochemical systems the useful and observable range for FRET measurements is on the order of 10-80Å.²³⁻²⁵

Energy transfer has been used for DNA recognition by a number of groups. Hélène and coworkers, for example, used hybridization to detect DNA by hybridization with two probes labeled with spectrally overlapping dyes.²⁶ Chin and coworkers have applied this type of system to the detection of hybridization within cells by microinjecting the fluorescent probes into cultured cells.²⁷ More recently, a powerful variation of this technique has been used by Tyagi and coworkers.²⁸ These so called “molecular beacons” hybridize to complementary DNA and result in a gain of signal through relief of quenching. The probe oligonucleotide forms a hairpin structure that brings a donor and quencher dye within close proximity. Upon hybridization to a complementary sequence, the donor and

quencher are separated by sufficient distance to allow for efficient fluorescence. This approach has proved to be quite powerful for genotyping and other applications.^{29,30} The gain-of-signal approach results in high levels of sensitivity. For detection of DNA in homogenous systems, this technique is clearly one of the best developed to date.

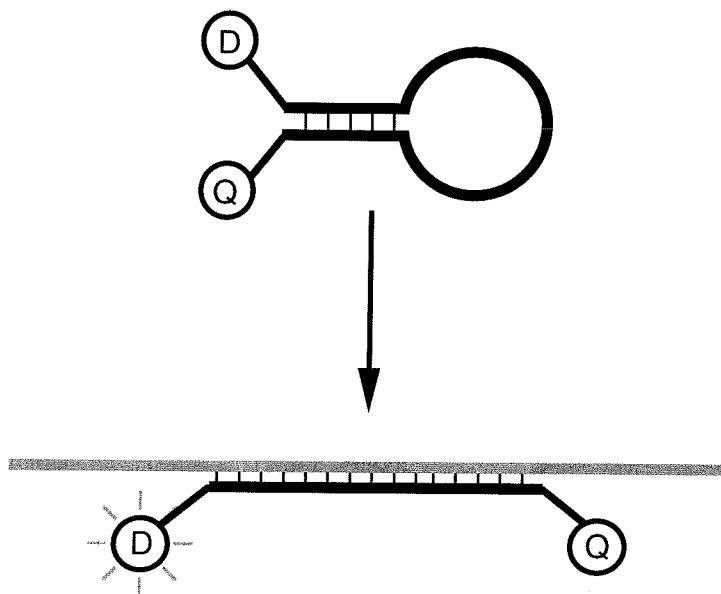


Figure 16. "Molecular beacon" approach to DNA detection.

We sought to develop a polyamide based system for DNA detection through FRET. A simple scheme for polyamide detection using FRET is outlined in figure 14. In this scheme, two polyamide-dye conjugates bind to adjacent binding sites. The dyes are chosen such that irradiation of the donor dye does not significantly excite the acceptor. The dyes should also possess significant spectral overlap between the emission band of the donor dye with the absorbance band of the acceptor. A significant advantage of this approach over other polyamide based approaches is that two compounds are required for the signal generation. For sequence specific applications, this could push the specificity to levels appropriate for the detection of a single position in a genome.

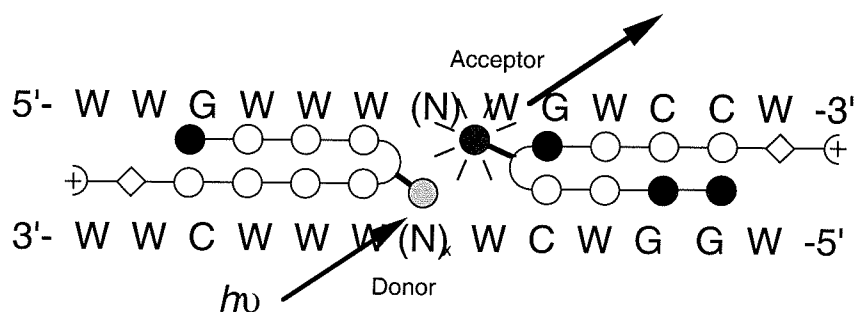


Figure 17. Polyamide detection of DNA using FRET.

An initial system was designed to test this scheme. Specifically, relatively small fluorescent dyes were chosen that met the above criteria. Specifically, a coumarin derivative was chosen as a donor because of its large Stoke's shift. The acceptor chosen was BODIPY FI, a small stable fluorescein analog. While this system did not show energy transfer in initial experiments, the approach may still be valid.

2.3.2 Design

The system was designed using two hairpin polyamides that bind with high specificity and affinity. The compounds chosen, ImPyPyPy- γ -PyPyPyPy- β -Dp¹⁷ and ImImPyPy- γ -ImPyPyPy- β -Dp,²² also bind to sites that are double mismatches for each other. Both parent compounds also bind DNA in an oriented fashion. This design was chosen such that the spacing of the donor and acceptor dyes could be controlled. It was chosen to append the dyes to the hairpin turn moiety, through the use of the (R)- α,γ -diaminobutyric acid residue. This design allows for a more rigid linkage between the polyamide and the dye molecule.

Dyes were chosen that met the spectral requirements for energy transfer to be observed. The coumarin compound has a large Stoke's shift, allowing irradiation far from the flourescein analog's absorbance band. The emission of the coumarin, however, shows significant overlap with the absorbance of the acceptor. The BODIPY-FI analog was

chosen over fluorescein because of its small size and superior photostability. Otherwise, this system is similar to that used by Tsien and coworkers for recent biological applications.³¹

2.3.3 Synthesis of Compounds

The conjugates used in these experiments were synthesized by standard solid phase synthesis incorporating an Fmoc protected (R)- α,γ -diaminobutyric acid residue. The Fmoc group was removed prior to cleavage of the polyamides from the resin. The polyamide primary amines were HPLC purified and readily coupled with commercially available succinimidyl esters of the dyes.

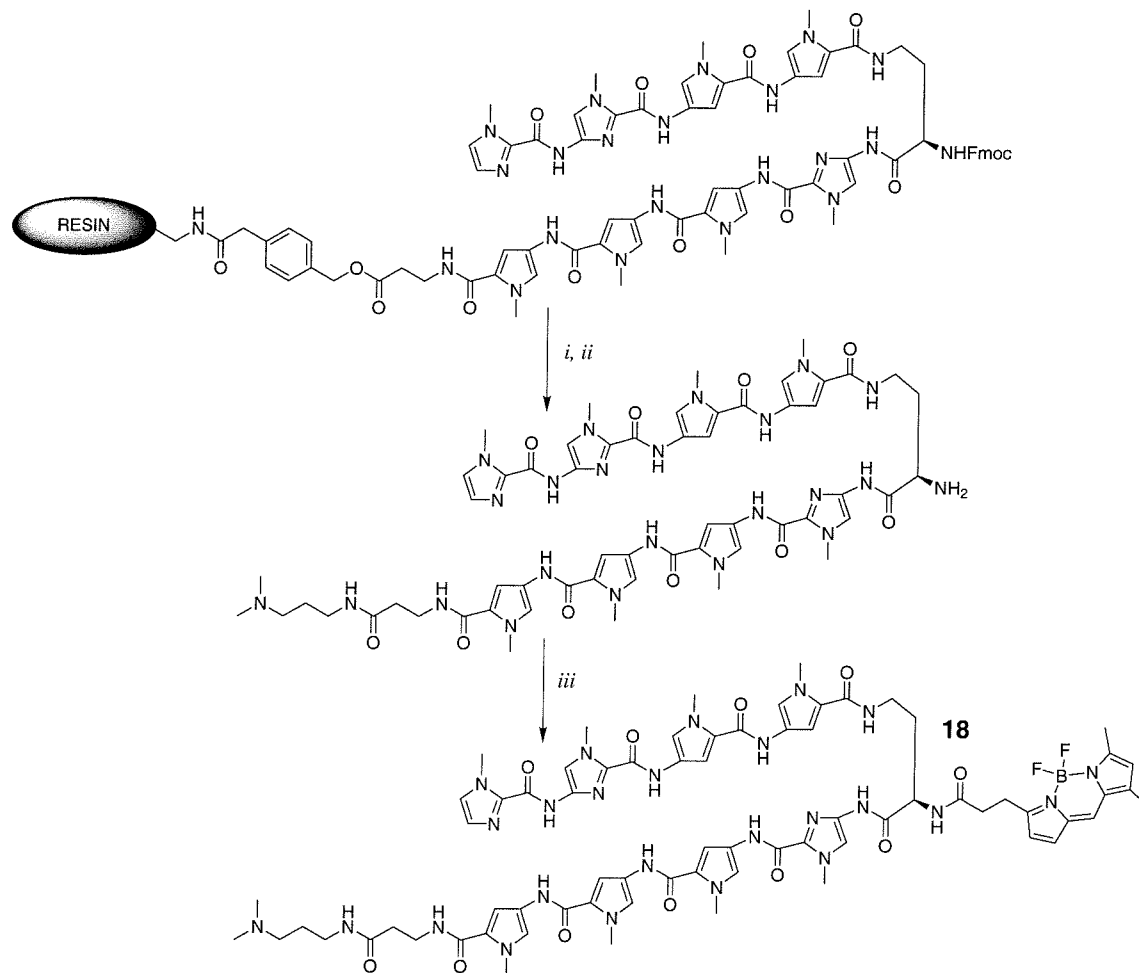


Figure 18. Synthesis of ImImPyPy- γ (BODIPY-FI)-ImPyPyPy- β -Dp conjugate **18.** *i*) 20% piperidine/DMF; *ii*) dimethylaminopropylamine, 55 °C; *iii*) BODIPY-FI, NHS ester (Molecular Probes, Inc.).

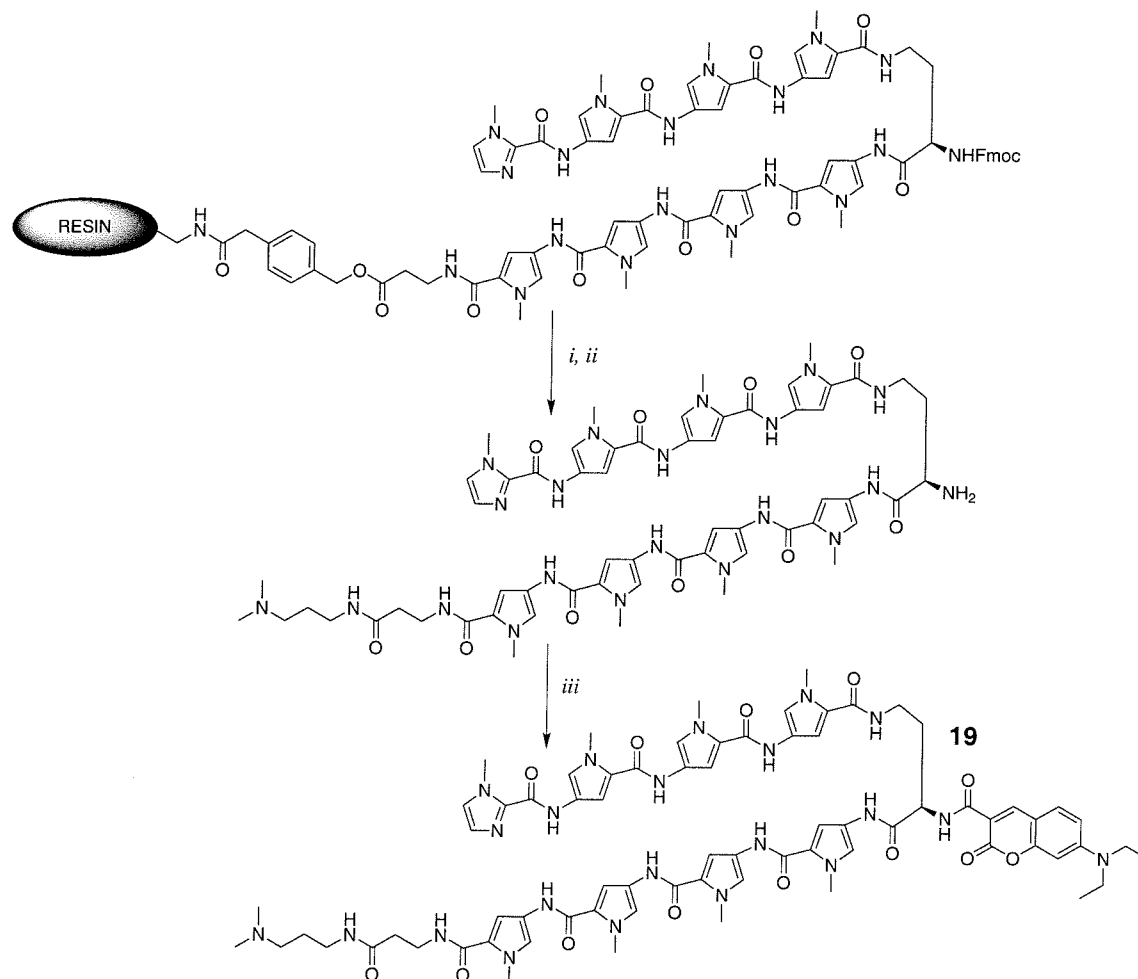


Figure 19. Synthesis of ImImPyPy- γ (coumarin)-ImPyPyPy- β -Dp conjugate **19.** *i*) 20% piperidine/DMF; *ii*) dimethylaminopropylamine, 55 °C; *iii*) 7-diethylaminocoumarin-3-carboxylic acid, *N*-hydroxysuccinimidyl ester (Molecular Probes, Inc.).

2.3.4 Results and Discussion

The designed polyamides were highly fluorescent. As designed, the polyamide coumarin conjugate **19** could be efficiently excited with 420 nm light, giving a 460 nm emission band (figure 20). The polyamide conjugate with BODIPY F1 was highly fluorescent when excited at 490 nm, but essentially non-fluorescent when excited at 420 nm. No effect on fluorescence was seen in the presence of DNA for either dye conjugate.

A series of oligonucleotides were designed with appropriate binding sites for polyamide conjugates **18** and **19** with intervening sequences ranging from 1 base pair to 9 base pairs. The fluorescence of the polyamide conjugates was monitored after equilibration with these oligoduplexes. Excitation of the complex with 420 nm light produced no increase in fluorescence of the BODIPY FI fluorophore as a function of either distance or the presence of DNA.

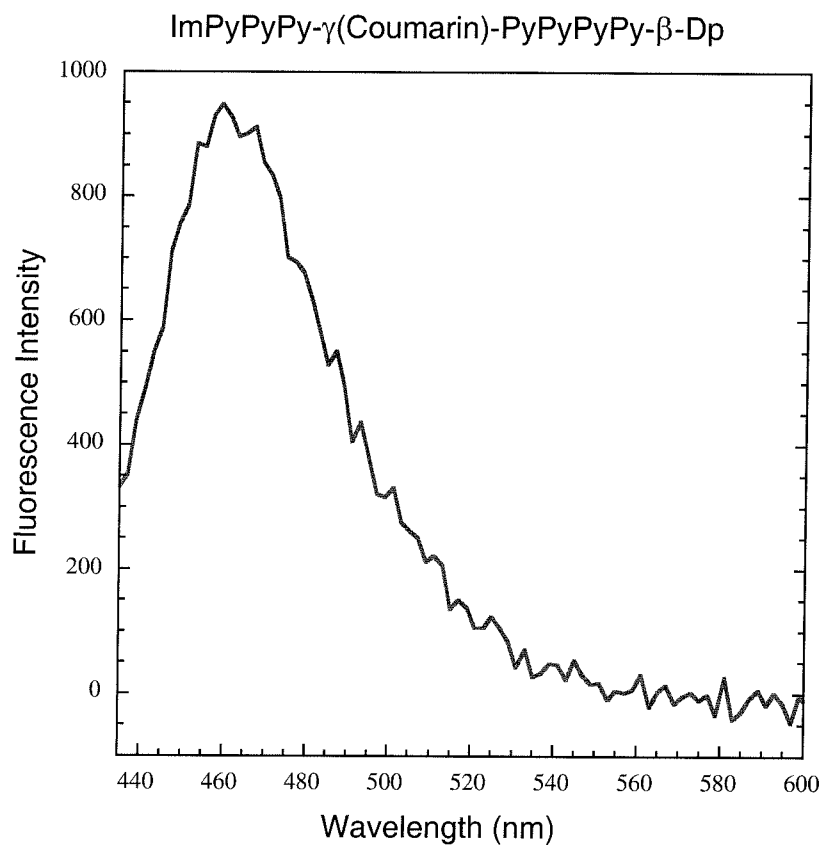


Figure 20. Fluorescence of coumarin conjugate 19. Fluorescence emission spectrum of a 100 nM solution was measured in water with excitation at 420 nm in a 0.5 cm cuvette. Slit widths were set to 5 nm. Data were collected at 2 nm increments with 1 sec integration.

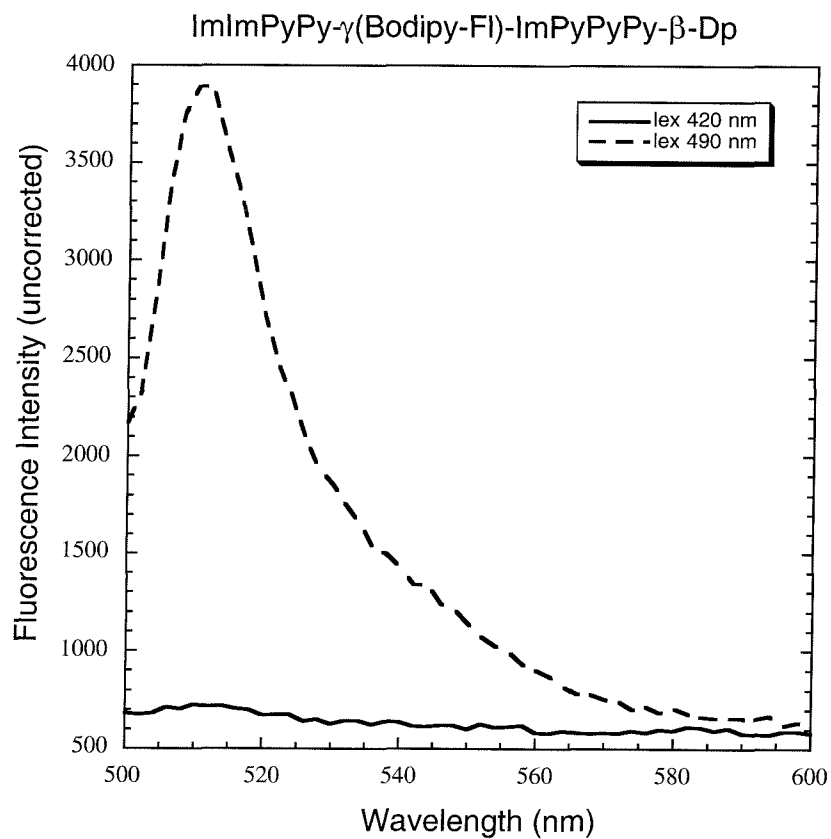


Figure 21. Fluorescence BODIPY-FI conjugate 18. Fluorescence emission spectrum of a 100 nM solution was measured in water with excitation at 420 nm (solid line) or 490 nm (dashed line) in a 0.5 cm cuvette. Slit widths were set to 5 nm. Data were collected at 2 nm increments with 1 sec integration.

ET-1

5'-GCC GGT **TGT ATA TTG ACC** ATG CGC C-3'
 3'-CGG CCA **ACA TAT AAC TGG** TAC GCG G-5'

ET-2

5'-GCC GGT **TGT ATA** TAT **TGA CCA** TGC GCC-3'
 3'-CGG CCA **ACA TAT** ATA **ACT GGT** ACG CGG-5'

ET-3

5'-GCC GGT **TGT ATA** TAT **ATT GAC CAT** GCG CC-3'
 3'-CGG CCA **ACA TAT** ATA **TAA CTG GTA** CCG GG-5'

ET-4

5'-GCC GGT **TGT ATA** TAT ATA **TTG ACC** ATG CGC C-3'
 3'-CGG CCA **ACA TAT** ATA TAT **AAC TGG** TAC GCG G-5'

ET-5

5'-GCC GGT **TGT ATA** TAT ATA TAT **TGA CCA** TGC GCC-3'
 3'-CGG CCA **ACA TAT** ATA TAT ATA **ACT GGT** ACG CGG-5'

Figure 22. Oligoduplexes used in these experiments. Bold text indicates polyamide binding sites. Grey text indicates intervening sequence. Polyamide binding site spacings of 1, 3, 5, 7, and 9 nucleotides were studied.

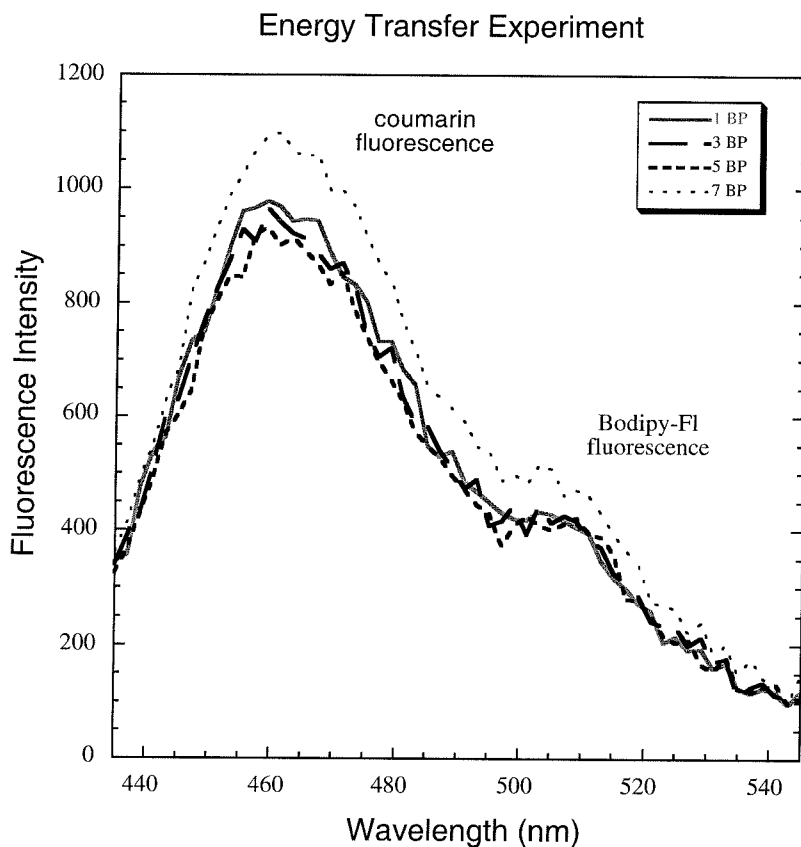


Figure 23. Energy transfer experiment. Polyamide dye conjugates **18** and **19** were mixed with oligoduplexes E1, E2, E3, or E4 each at 100 nM concentration. No energy transfer is observed relative to a solution containing no DNA (data not shown). The complex was excited at 420 nm in these experiments. Solution conditions were as follows: 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂ at pH 7.0.

2.3.5 DNA Binding of Polyamide-Fluorophore Conjugates

The ability of the compounds used in this study to bind to DNA sequence specifically was assayed using quantitative DNase I footprint titrations on plasmids containing appropriate match and mismatch binding sites. Both compounds bound to their designed binding sites with reduced affinity as compared to the parent hairpin polyamides.

The observed specificities were also somewhat reduced. Preliminary results of binding studies are summarized in table 3.

Table 3. Apparent equilibrium association constants (M^{-1}) of dye labeled polyamides.^a

Polyamide Conjugate	5'-TGGTCA-3'	5'-TGTACA-3'
ImImPyPy-DABA-ImPyPyPy- β -Dp	1.1×10^9	2.0×10^8
ImImPyPy- γ (BODIPY-Fl)-ImPyPyPy- β -Dp (18)	9.8×10^8	3.8×10^8
	5'-ttAGTATTTtg-3'	5'-ttAGTACTtg-3'
ImPyPyPy- γ -PyPyPyPy- β -Dp ¹⁷	3.5×10^9	5.0×10^8
ImPyPyPy- γ (Coumarin)-PyPyPyPy- β -Dp (19)	1.4×10^8	4.0×10^7

^a Quantitative DNase I footprinting reactions were carried out on a 3'-labeled EcoRI/PvuII restriction fragment derived from plasmid pJT8 as described.¹⁷ Data represent single trials. Data for compound **18** were collected in collaboration with Ralf Jäger.

2.3.6 Energy Transfer Conclusions

Several reasons are possible for the failure of this system to generate an energy transfer dependent acceptor fluorescence. One possibility for the failure of this system to exhibit pyrene excimer fluorescence lies in the reduced affinity of coumarin derivative **19**. At the concentrations used in the fluorescence experiments, the targeted binding site should have been saturated; however, given the reduced specificity of the conjugate, it is possible that compound did not fully saturate the match site. The BODIPY-Fl conjugate showed little difference from the parent hairpin containing an underivatized diaminobutyric acid turn. While the results shown are from a single experiment, the compounds were run side by side, and were also qualitatively similar. It is expected that use of a coumarin derivative with a longer linker might prove advantageous.

A second possibility exists for the failure of this experiment. The overlap between the donor and acceptor may not be sufficient for efficient energy transfer. As FRET is dependent on the quantum yield of the donor dye, a brighter donor dye might be more effective. A more thorough study of energy transfer with a greater variety of dye conjugates should be undertaken.

An additional benefit of this study is that these dye-labeled compounds are also amenable to the study of the cell permeability of polyamides. The coumarin dye used in this study is quite similar to one used recently by Tsien and coworkers for an *in vivo* fluorescence experiment. Both dyes are relatively small, and have both large extinction coefficients and quantum yields. Use of fluorescent conjugates may provide information regarding not only the membrane permeation of these compounds, but also of their subcellular localization. These compounds are being assayed for their ability to permeate mammalian cells in cell culture experiments.

2.4 Detection Through Pyrene Excimer Formation

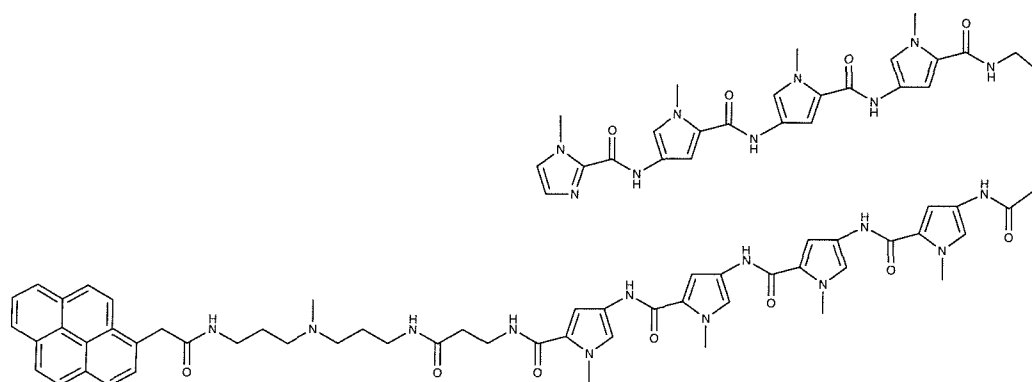
2.4.1 Background

A second energy transfer phenomenon that has been put to use in biochemical assays is that of pyrene excimer formation. The planar aromatic pyrene shows a concentration dependent fluorescence emission spectra. The high concentration spectra results from the formation of an excited state dimer resulting from close physical proximity of two dye molecules. This effect can be produced at lower concentrations by use of a covalent tether. Bringing two pyrene molecules together has been used for the study of membrane dynamics, as well as other biological processes. A group has recently reported the use of pyrene excimer formation to report DNA hybridization.³² Two oligonucleotides were labeled with pyrene. Upon hybridizing to a targeted binding site on single stranded DNA, the pyrenes were forced into close proximity, and excimer fluorescence was induced. We sought a similar result through the introduction of pyrenes on polyamide molecules.

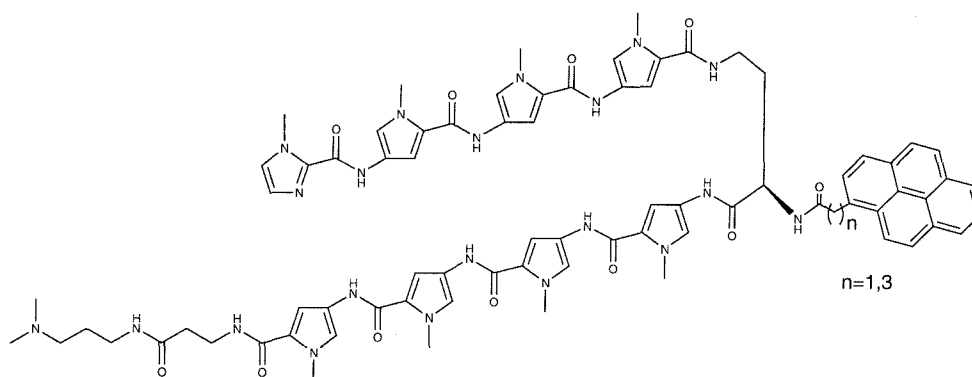
2.4.2 Design of Pyrene Conjugates

Computer modeling was used to design possible binding modes of dye labeled polyamides. For formation of pyrene excimers, two binding models were employed: one with hairpins held in a head to tail fashion, the second in a turn to turn fashion. These binding models are shown in figure 25. They were generated with the program InsightII 97.0 (MSI, Carlsbad, CA) using simple minimization and the CVFF forcefield. The models were used only as a rough guide to the desired linker lengths, and DNA binding site separations of the polyamides. Using these models as guidelines, three pyrene conjugates were designed with the structures shown in figure 24. According to these

models, the turn to turn mode allows for the polyamides to be held at closer proximity than that of the tail to tail binding mode.



20 ImPyPyPy- γ -PyPyPyPy- β -Dp-AcPYR



21 $n=1$ ImPyPyPy- γ (AcPYR)-PyPyPyPy- β -Dp
 22 $n=3$ ImPyPyPy- γ (BuPYR)-PyPyPyPy- β -Dp

Figure 24. Pyrene conjugates used in this study.

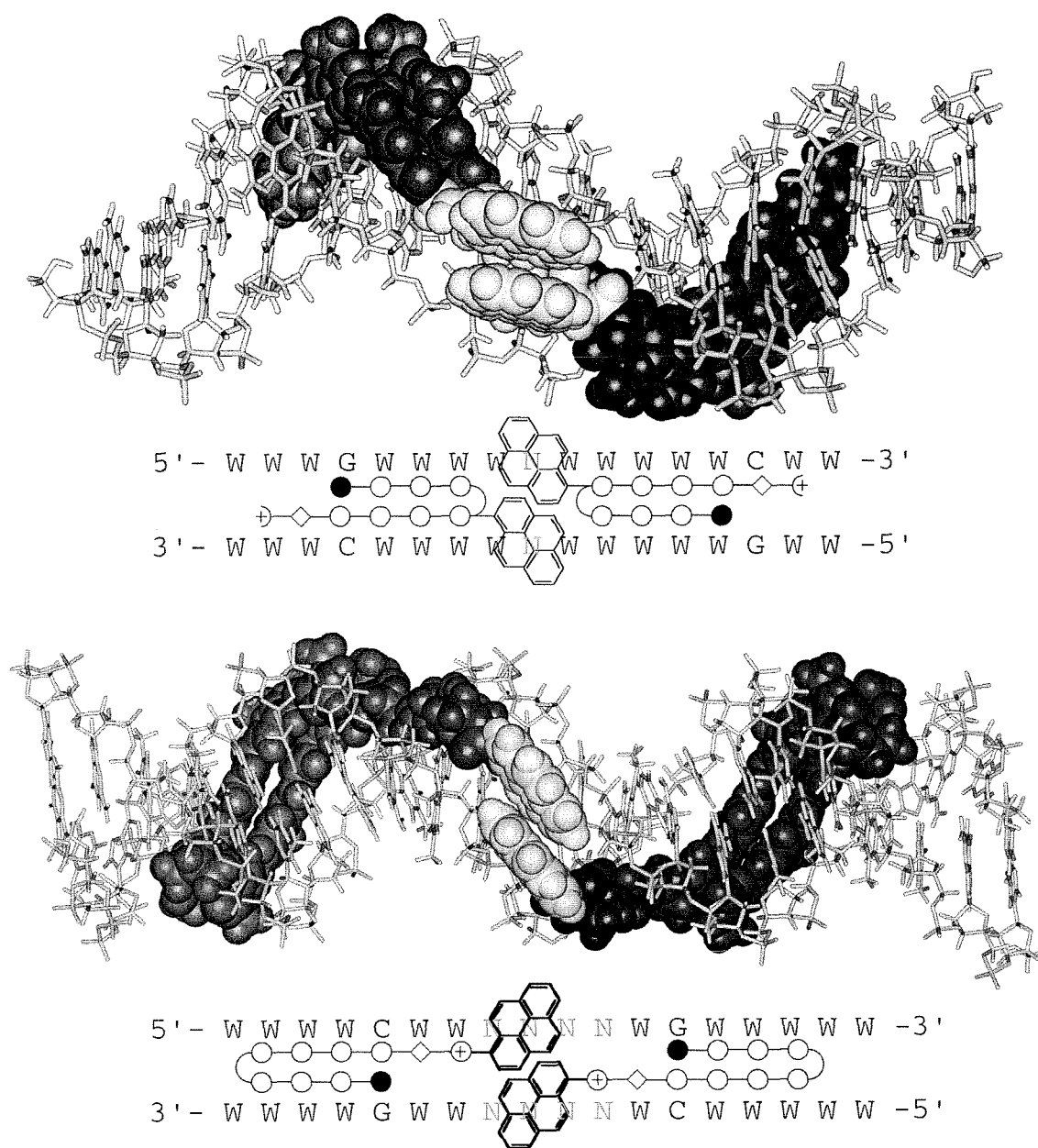


Figure 25. Computer model of pyrene polyamide binding modes.

Turn-to-Turn Dimerization:

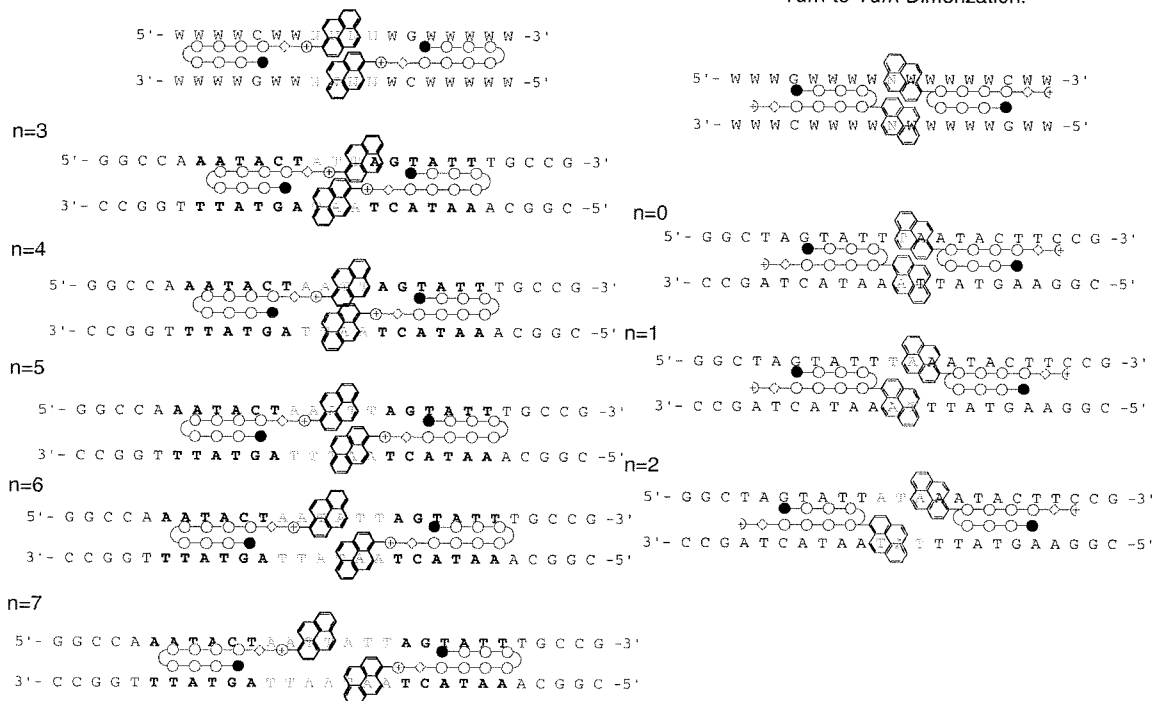


Figure 26. Pyrene oligos used for pyrene excimer experiments.

2.4.3 Results and Discussion

The pyrene labeled polyamides were readily synthesized using standard solid phase polyamide synthesis. Coupling of the pyrene to the diaminobutyric acid turn moiety was performed on the resin after removing the Fmoc protecting group. The C terminally labeled compound was synthesized according to the same strategy outlined in figure 4 for compound **14**. The compounds were purified by reversed phase HPLC, and characterized by MALDI-TOF mass spectrometry.

These compounds exhibited the anticipated pyrene monomer fluorescence spectra in the absence of DNA. Figure 30 shows the spectra of compound **21** in the presence and absence of DNA sequences designed to align the pyrenes. The structured fluorescence bands from 370 to 420 nm are typical for pyrene monomer fluorescence. The expected

excimer fluorescence would occur as a broad unstructured peak at roughly 500 nm. Spectra were collected both under air, and in degassed solutions under Ar with little difference (data not shown) except in reducing the broad fluorescence band at 440 nm. Data for all three compounds resembled that presented for compound **21**.

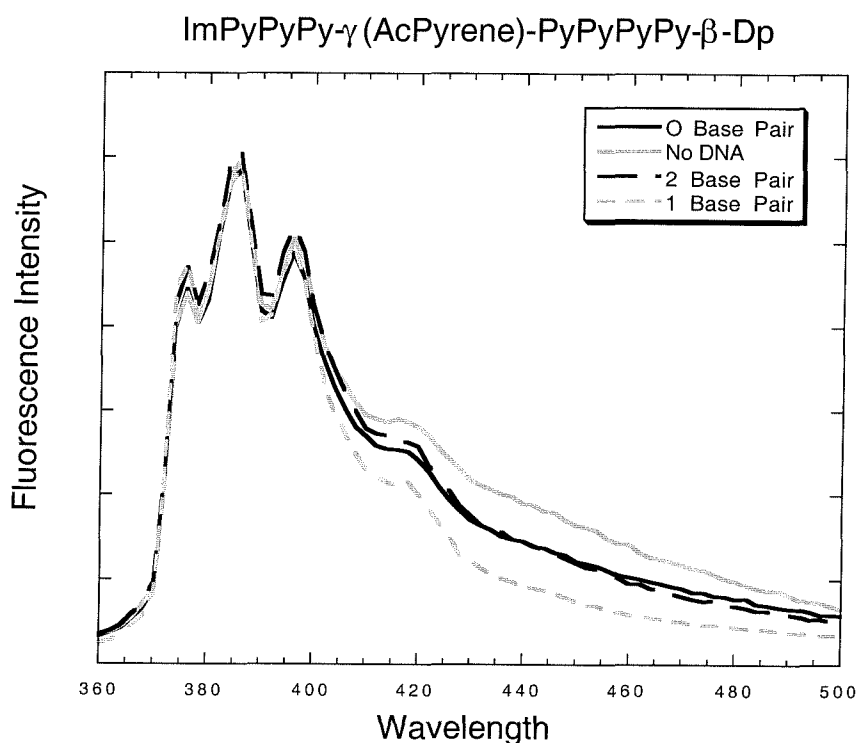


Figure 27. Fluorescence spectra of turn to turn pyrene system. 400 nM **21** was equilibrated with either no DNA or 200 nM of the appropriate duplex. Spectra were obtained with $\lambda_{\text{ex}} = 345$ nm in 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, pH 7.0.

The failure of these systems to display pyrene excimer formation was unexpected. As the DNA binding modes of the parent hairpin compounds were well characterized, the distance constraints for the pyrene-pyrene distances seem well within those found in

successful system. Some insight into this failure can be gleaned from the analysis of excimer formation in pyrene crystals.³³ The excimer phenomenon shows some requirement for stacking of the pyrenes. This can be achieved through covalent rigid linkage or through enough flexibility to achieve this conformation during the relatively long half-life of the excited state.

Pyrene is known to interact weakly with DNA through both intercalative and groove binding modes.^{34,35} The effect of delivering the pyrene moiety to the DNA with the polyamide may have facilitated this interaction. With the effective concentration of the pyrene thus increased, it may adopt a stable conformation either wholly or partially intercalated. This proposed interaction could preclude the interaction of the two pyrene moieties delivered to the adjacent binding sites on the DNA target.

Table 4. Apparent equilibrium association constants of pyrene conjugates.^a

Polyamide Conjugate	5'-ttAGTATTtg-3'	5'-ttAGTACTig-3'
ImPyPyPy- γ -PyPyPyPy- β -Dp ¹	3.5×10^9	5.0×10^8
ImPyPyPy- γ -PyPyPyPy- β -Dp-AcPyrene (20)	1.3×10^9	7.7×10^8
ImPyPyPy- γ (AcPyrene)-PyPyPyPy- β -Dp (21)	$< 1 \times 10^7$	$< 1 \times 10^7$
ImPyPyPy- γ (BuPyrene)-PyPyPyPy- β -Dp (22)	$< 1 \times 10^7$	$< 1 \times 10^7$

^a Quantitative DNase I footprinting reactions were carried out on a 3'-labeled EcoRI/PvuII restriction fragment derived from plasmid pJT8 as described.¹⁷ Data represent single trials.

This hypothesis is substantiated for compound **20** which shows markedly reduced specificity from the parent hairpin polyamide. As shown in table 4, conjugate **20**, binds to a 5'-AGTATT-3' match site with an affinity comparable to that of the parent hairpin polyamide. The mismatch also binds more tightly, as do all other binding sites on the gel. The compound is almost completely non-specific at 100 nM and above, the concentration range used for the fluorescence experiments.

Unlike the tail linked compound **20**, the compounds with pyrene attached on the gaba turn, **21** and **22**, show markedly reduced affinities and specificities. Overall, it seems that linkage of dyes at the turn residue gives mixed results, particularly for

compounds that interact directly with the DNA. It is possible that such conjugates cause the gaba turn to adopt an incompatible conformation that precludes high affinity binding. Conjugates with dyes not known to interact with DNA, such as BODIPY-Fl described above, may not cause such distortion in the bound complex.

2.5 Conclusions and Future Outlook

Polyamide conjugates with fluorescent dyes promise to be useful and interesting molecules. Of the conjugates described, the conjugate with BODIPY-Fl showed the least perturbation of the properties of the parent hairpin polyamide. Conjugation of such dyes with polyamides with higher affinities and specificities could result in useful materials for the study of DNA both *in vitro* and *in vivo*. Several possible uses of such conjugates are outlined below.

2.5.1 Genomic Analysis

Polyamides can be used to deliver fluorescent molecules to double stranded DNA. In particular, conjugates prepared with cyanine dyes show promising degrees of sequence selectivity, fluorescence enhancement, and affinity. Detection of longer DNA sequences than is currently practical with the polyamide approach will be necessary for polyamide conjugates to become useful reagents for sequence-specific genomic analysis. The approach outlined by Castro¹³ could also be applied using polyamides. In such a system DNA fragments are hybridized to fluorescent probes and flowed past a detector a single molecule at a time. Using two 15-mer probes was sufficient to detect a sequence in genomic DNA. Polyamides are beginning to approach this size of binding site. The system could possibly be extended to three or four fluorescently labeled polyamides. Also, a similar experiment might be possible with capillary electrophoresis. As the binding site

sizes of polyamides are increased, such conjugates might also find use in direct optical mapping techniques.

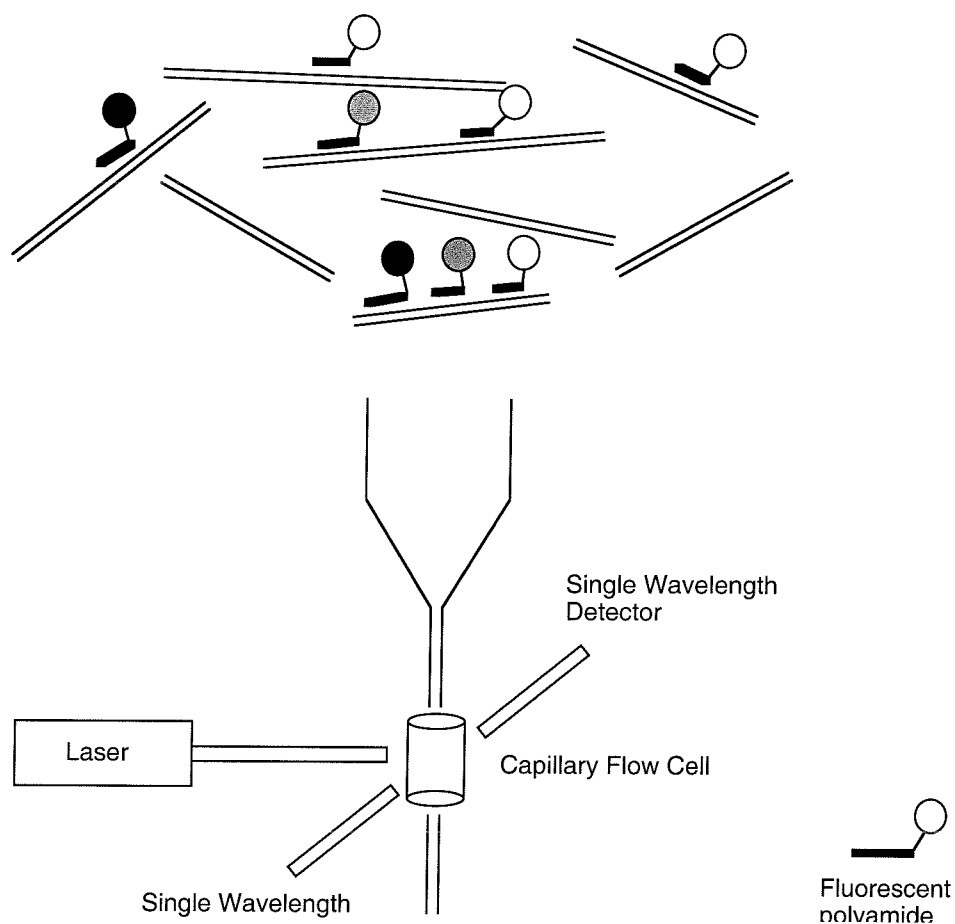


Figure 28. Detection of DNA by coincidence detection. Single molecules pass detector. Sequences are detected by simultaneous binding of 2 or more polyamides to a single molecule.

2.5.2 Biophysical Analysis

Fluorescent polyamide conjugates may be useful for thorough biochemical analysis of the thermodynamic and kinetic properties of polyamide binding to DNA. The cyanine dye conjugates, for instance, could be used to accurately measure the association rate for DNA binding by direct fluorometric observation. Likewise, kinetics could be measured by

labeling a synthetic oligoduplex with a fluorophore and using FRET to monitor the binding of a fluorophore labeled polyamide to the DNA. BODIPY conjugate **18** is well suited to such measurements.

As evidence also suggests that the cyanine dye conjugates bind intercalatively, these may be interesting compounds for disrupting major groove binding proteins' interactions with DNA. Such intercalative complexes have been investigated by R. Bremer, but second generation compounds could prove useful. It is still a major goal to extend the ability of polyamides to disrupt protein DNA interactions in both the major and minor grooves.

2.5.3 Cellular Studies

Fluorescent dyes are useful probes at the cellular and subcellular level. There are fundamental questions remaining regarding a polyamide's fate in a cell. Developing cell permeable polyamide dye conjugates would enable direct observation of cell permeation by polyamides. Several of the dye conjugates described in this thesis could be useful for such applications. The cyanine dyes, for instance, normally do not permeate cells. Small modifications could potentially produce cell permeable derivatives. The small dyes used including coumarin and the BODIPY family are well suited to such biological studies. Such properties are difficult to predict except by experimentation. Such cell permeable conjugates could be useful for applications like *in situ* fluorescence detection of DNA sequences or karyotyping.

2.6 Experimental

Activated benzothiazole (1). 14.9 g (0.1 mol) of 2-methylbenzothiazole was heated to 100 °C in the presence of 3 eq. of methyl iodide (41.4 g, 0.3 mol) in 100 mL of dioxanes. After 2 hours, the reaction was cooled and precipitated with ethyl ether. The off-white solid was collected by filtration to yield 27.2 g (94%) of 1,2-dimethylbenzothiazole. 2.5 g of this product was heated to reflux in acetic anhydride in the presence of excess diphenylimine (10 g, 51 mmol, 6 eq.) After one hour, the solution was partitioned between dichloromethane and brine. The product was precipitated from the dichloromethane layer with ethyl ether to yield 1.5 g (3.4 mmol, 40%). ¹H NMR (DMSO-*d*₆) δ 2.04 (s, 3H), 3.85 (s, 3H), 5.64 (d, 1H, *J*=13.9 Hz), 7.5-7.7 (m, 6H), 8.07 (d, 1H, *J*=8.2 Hz), 8.34 (d, 1H, *J*=7.9 Hz), 8.75 (d, 1H, *J*=13.9 Hz).

Ethyl lepidine butyrate bromide (5). 10 g (70 mmol) of lepidine and 13.7 g (70 mmol) of ethyl-4-bromobutyrate were heated together at 120 °C for 1 hour. Upon cooling, the solution crystallizes to give a solid mass. The solid was crushed, triturated with acetone, and dried *in vacuo* to give 23.1 g of **5**. ¹H NMR (DMSO-*d*₆) δ 1.12 (t, 3H, *J*=7.1 Hz), 2.20 (m, 2H), 2.58 (t, 2H, *J*=7.3 Hz), 3.00 (s, 3H), 3.97 (q, 2H, *J*=7.1 Hz), 5.11 (t, 2H, *J*=7.6 Hz), 8.04 (t, 1H, *J*=7.5 Hz), 8.11 (d, 1H, *J*=6.0 Hz), 8.26 (t, 1H, *J*=7.5 Hz), 8.53 (d, 1H, *J*=8.4 Hz), 8.69 (d, 1H, *J*=9.0 Hz), 9.56 (d, 1H, *J*=6.1 Hz). ¹³C NMR (DMSO-*d*₆) δ 13.9, 19.7, 24.6, 30.2, 55.9, 60.0, 119.2, 122.6, 127.0, 128.9, 129.5, 135.0, 136.7, 148.4, 158.6, 171.9.

2-thiomethyl-3-methylbenzothiazole iodide (6). According to the general procedure of Benson *et al.*,⁵ 20 g (110 mmol) of 3-methyl benzothiazole-2-thione were refluxed overnight in 700 mL of absolute ethanol in the presence of 50 g (352 mmol, 3.2

eq.) of methyl iodide. The solution clarified after refluxing. Upon cooling, ethyl ether was added until a white precipitate formed. The product was collected by filtration, washed with ethyl ether, and dried *in vacuo*, giving 32.7 g (101 mmol, 92%) of pure product. ^1H NMR ($\text{DMSO-}d_6$) δ 3.12 (s, 3H), 4.10 (s, 3H), 7.71 (t, 1H), 7.83 (t, 1H), 8.18 (d, 1H, $J=8.4$ Hz), 8.40 (d, 1H, $J=8.1$ Hz).

Ethyl 4-picoline butyrate bromide (7). 9.3 g (100 mmol) of 4-picoline and 19.5 g (100 mmol) of ethyl-4-bromobutyrate were heated together at 120 °C for 1 hour. Upon cooling, the solution crystallizes to give a thick oil. The oil was dried *in vacuo* to give 27.6 g (96%) of **7** as a yellow white solid. ^1H NMR ($\text{DMSO-}d_6$) δ 1.12 (t, 3H, $J=7.1$ Hz), 2.13 (m, 2H), 2.4 (t, 2H, $J=7.4$ Hz), 2.59 (s, 3H), 3.98 (q, 2H, $J=7.1$ Hz), 4.62 (t, 2H, $J=7.1$ Hz), 8.00 (d, 1H, $J=6.3$ Hz), 9.01 (d, 1H, $J=6.7$ Hz) ^{13}C NMR ($\text{DMSO-}d_6$) δ 14.1, 21.4, 26.0, 30.1, 59.0, 61.2, 128.4, 143.9, 159.0, 171.8.

Thiazole blue butyric acid ethyl ester (8). 676 mg (2 mmol) lepidine salt **5** and 618 mg of **1** (2 mmol) were refluxed in absolute ethanol (20 mL). Upon addition of triethylamine (280 μL , 2 mmol), the solution turned a dark blue color. Refluxing was continued for 10 minutes. The solution was cooled to room temperature and precipitated with ethyl ether to give thiazole orange ethyl ester **8**, 0.98 g (~92% assuming iodo salt) as an orange red solid. ^1H NMR ($\text{DMSO-}d_6$) δ 1.14 (t, 3H, $J=7.1$ Hz), 2.10 (m, 2H), 2.45 (m, 2H), 3.73 (s, 3H), 4.00 (q, 2H, $J=7.1$ Hz), 4.53 (t, 2H, $J=7.1$ Hz), 6.50 (d, 1H, $J=12.4$ Hz), 7.08 (d, 1H, $J=13.3$ Hz), 7.30 (t, 1H, $J=7.8$ Hz), 7.48 (t, 1H, $J=7.4$ Hz), 7.58 (d, 1H, $J=8.2$ Hz), 7.68 (m, 2H), 7.82 (d, 1H, $J=7.2$ Hz), 7.87 (d, 1H, $J=7.6$ Hz), 7.94 (t, 1H, $J=7.5$ Hz), 8.15 (t, 1H, $J=12.7$ Hz), 8.33 (d, 1H, $J=7.1$ Hz), 8.44 (d, 1H, $J=8.3$ Hz).

Thiazole blue butyric acid (9). 200 mg of **8** were stirred in 10 mL of methanol with 10 mL of 1M KOH at room temperature for 4 hours. The reaction was neutralized with 1M HCl and applied to a reversed phase column equilibrated with 0.1% trifluoroacetic acid. The column was washed with water, and the acid eluted with 25% acetonitrile, 0.1% trifluoroacetic acid to yield thiazole blue ethyl ester **9** as an dark blue solid.

Thiazole orange butyric acid ethyl ester (10). 676 mg (2mmol) lepidine salt **5** and 646 mg of **6** were refluxed in absolute ethanol (20 mL). Upon addition of triethylamine (280 μ L, 2 mmol) the solution turned a deep red color. Refluxing was continued for 10 minutes. The solution was cooled to room temperature and precipitated with ethyl ether to give thiazole orange ethyl ester **10**, 0.98 g (~92% assuming iodo salt) as an orange red solid. ^1H NMR (DMSO- d_6) δ 1.17 (t, 3H, $J=1.17$ Hz), 2.11 (m, 2H), 4.02 (s, 3H), 4.02 (q, 2H, $J=7.1$ Hz), 4.60 (t, 2H, $J=7.0$ Hz), 6.90 (s, 1H), 7.32 (d, 1H, $J=7.2$ Hz), 7.42 (t, 1H, $J=7.7$ Hz), 7.60 (t, 1H), 7.75 (t, 2H, $J=5.9$ Hz), 8.03 (m, 2H), 8.17 (d, 1H, $J=8.8$ Hz), 8.59 (d, 1H, $J=7.2$ Hz), 8.78 (d, 1H, $J=8.6$ Hz). ^{13}C NMR (DMSO- d_6) δ 14.1, 24.1, 30.3, 30.7, 53.4, 60.2, 88.2, 107.9, 113.0, 118.0, 122.9, 123.9, 124.2, 124.5, 125.9, 126.8, 128.2, 133.3, 137.1, 140.4, 144.3, 148.5, 160.1, 172.2.

Thiazole orange butyric acid (11). 53 mg (0.1 mmol) ester was stirred in 2 mL of water with 2 mL of 1M KOH. The reaction was heated briefly to 60 $^{\circ}\text{C}$ and cooled to room temperature. The basic solution was neutralized with 1M HCl, applied to a small column of reversed phase silica, and washed with 0.1% trifluoroacetic acid (TFA) in water. Pure acid was eluted from the column with 25% acetonitrile in 0.1% TFA. The collected fractions were lyophilized to give 40 mg of **11** (0.08 mmol, 85%). ^1H NMR (DMSO- d_6) δ 2.11 (m, 2H), 2.48 (m, 2H), 4.02 (s, 3H), 4.60 (t, 2H), 6.93 (s, 1H), 7.38 (m, 2H),

7.61 (t, 1H), 7.78 (m, 2H), 7.99 (t, 1H), 8.05 (d, 1H), 8.21 (d, 1H, $J=8.6$ Hz), 8.62 (d, 1H, $J=7.0$ Hz), 8.80 (d, 1H, $J=8.3$ Hz).

Thiazole yellow butyric acid ethyl ester (12). 2.89 g (10 mmol) picoline salt **7** and 3.23 g (10 mmol) of **6** were refluxed in absolute ethanol (20 mL). Upon addition of triethylamine (1.4 mL, 10 mmol), the solution turned a dark yellow brown color. Refluxing was continued for 10 minutes. The solution was cooled to room temperature and precipitated with ethyl ether to give 2.78 g of thiazole orange ethyl ester **10** (~58% assuming iodo salt). ^1H NMR (DMSO- d_6) δ 1.16 (t, 3H, $J=1.17$ Hz), 2.08 (m, 2H), 2.38 (m, 2H), 3.73 (s, 3H), 4.01 (q, 2H, $J=7.1$ Hz), 4.24 (t, 2H, $J=6.9$ Hz), 6.27 (s, 1H), 7.30 (t, 1H, $J=7.5$ Hz), 7.39 (d, 2H, $J=6.7$ Hz), 7.52 (t, 1H, $J=7.3$ Hz), 7.60 (d, 1H, $J=8.1$ Hz), 7.91 (d, 1H, $J=7.8$ Hz), 8.33 (d, 2H, $J=6.8$ Hz). ^{13}C NMR (DMSO- d_6) δ 14.0, 18.5, 25.6, 30.1, 32.9, 56.7, 60.0, 112.0, 118.3, 122.5, 123.3, 123.4, 127.7, 140.3, 141.2, 150.3, 157.1, 171.8.

Thiazole yellow butyric acid (13). 53 mg (0.1 mmol) ester was stirred in 2 mL of water with 2 mL of 1M KOH. The reaction was heated briefly to 60 °C and cooled to room temperature. The basic solution was neutralized with 1M HCl, applied to a small column of reversed phase silica, and washed with 0.1% trifluoroacetic acid (TFA) in water. Pure acid was eluted from the column with 25% acetonitrile in 0.1% TFA. The collected fractions were lyophilized to give 40 mg of **11** (0.08 mmol, 85%). ^1H NMR (DMSO- d_6) δ 2.04 (m, 2H), 2.30 (t, 2H, $J=7.1$ Hz), 3.72 (s, 3H), 4.23 (t, 2H, $J=7.1$ Hz), 6.25 (s, 1H), 7.29 (t, 1H, $J=7.5$ Hz), 7.38 (d, 2H, $J=7.2$ Hz), 7.51 (t, 1H, $J=7.2$ Hz), 7.59 (d, 1H, $J=8.2$ Hz), 7.90 (d, 1H, $J=7.8$ Hz), 8.34 (d, 2H, $J=7.1$ Hz). ^{13}C NMR (DMSO- d_6) δ 25.7, 30.2, 32.8, 57.0, 89.6, 112.0, 118.4, 122.5, 123.3, 123.5, 127.8, 140.4, 141.3, 150.4, 157.2, 173.3.

ImPyPyPy- γ -PyPyPyPy- β -Dp-TO (14). A solution 5 mg of ImPyPyPy- γ -PyPyPyPy- β -Dp-propylamine in 100 μ L of DMF was treated with a solution of 5 mg of thiazole orange acid (**11**) that had been activated with HBTU (3 mg) in 100 mL DMF in the presence of 50 μ L of diisopropylethylamine. The solution was allowed to stand at room temperature for 12 hours, then was filtered through a 0.2 μ m nylon filter and purified by HPLC to give pure **14**. MALDI-TOF MS obsvd. (M+H)⁺ 1623.76 calcd. 1623.74. ¹H NMR (DMSO-d₆) is shown below.

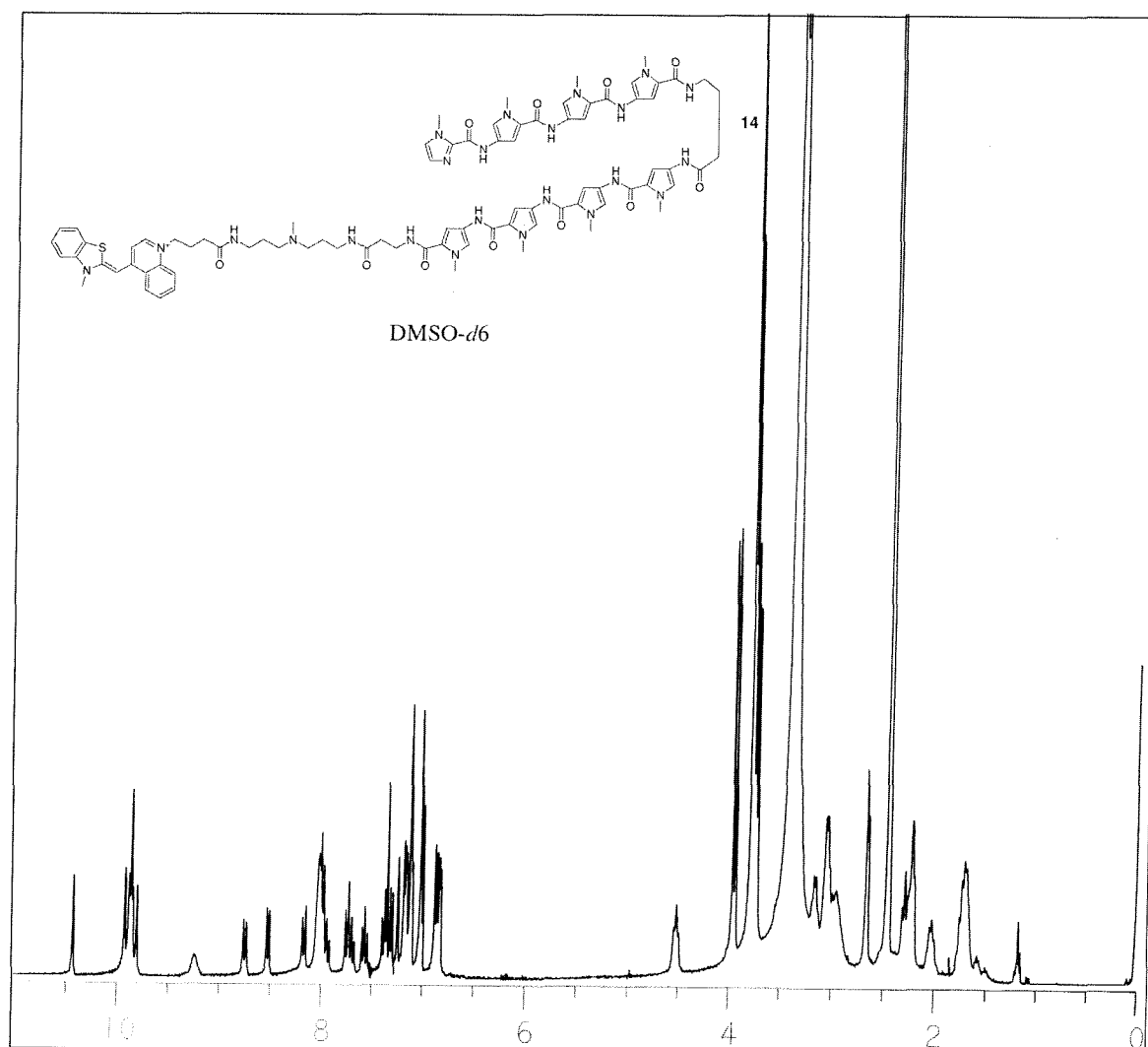


Figure 29. ¹H NMR of conjugate **14**.

ImPyPyPy- γ (TO)-PyPyPyPy- β -Dp (15). Thiazole yellow butyric acid 13 (5 mg) was activated with 3 mg of HBTU in 100 μ L DMF with 50 μ L of DIEA. This solution was added to a solution of 5 mg of ImPyPyPy-daba-PyPyPyPy- β -Dp in 100 mL DMF. The reaction was allowed to stand for 12 hours and then purified by reversed phase HPLC. Electrospray MS obsvd. (M+H)²⁺ 798.2 calcd. 798.35.

ImPyPyPy- γ (TY)-PyPyPyPy- β -Dp and ImPyPyPy- γ (TAB)-PyPyPyPy- β -Dp (16) and (17). Conjugates were synthesized as described for compound 15. **16:** Electrospray MS obsvd. (M+H)²⁺ 773.4 calcd. 773.35. **17:** Electrospray MS obsvd. (M+H)²⁺ 811.2 calcd. 811.36.

2.7 References

- (1) Levy, E. R.; Herrington, C. S. *Non-isotopic methods in molecular biology: a practical approach*; IRL Press: Oxford; New York, 1995.
- (2) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; 2nd Ed. ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 1989.
- (3) Brooker, L. G. S.; Keyes, G. H.; Williams, W. W. *J. Am. Chem. Soc.* **1942**, *67*, 1889-1893.
- (4) Rye, H. S.; Quesada, M. A.; Peck, K.; Mathies, R. A.; Glazer, A. N. *Nucl. Acids Res.* **1991**, *19*, 327-333.
- (5) Benson, S. C.; Singh, P.; Glazer, A. N. *Nucl. Acids Res.* **1993**, *21*, 5727-5735.
- (6) Spielmann, H. P.; Wemmer, D. E.; Jacobsen, J. P. *Biochem.* **1995**, *34*, 8542-8553.
- (7) Perkins, T. T.; Quake, S. R.; Smith, D. E.; Chu, S. *Science* **1994**, *264*, 822-826.
- (8) Gurrieri, S.; Wells, K. S.; Johnson, I. D.; Bustamante, C. *Anal. Biochem.* **1997**, *249*, 44-53.
- (9) Schwartz, D. C.; Samad, A. *Curr. Op. Biotech.* **1997**, *8*, 70-74.
- (10) Cai, W. W.; Jing, J. P.; Irvin, B.; Ohler, L.; Rose, E.; Shizuya, H.; Kim, U. J.; Simon, M.; Anantharaman, T.; Mishra, B.; Schwartz, D. C. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3390-3395.
- (11) Schwartz, D. C.; Li, X. J.; Hernandez, L. I.; Ramnarain, S. P.; Huff, E. J.; Wang, Y. K. *Science* **1993**, *262*, 110-114.
- (12) Nie, S.; Chiu, D. T.; Zare, R. N. *Anal. Chem.* **1995**, *67*, 2849-2857.
- (13) Castro, A.; Williams, J. G. K. *Anal. Chem.* **1997**, *69*, 3915-3920.
- (14) Benson, S. C.; Zeng, Z. X.; Glazer, A. N. *Anal. Biochem.* **1995**, *231*, 247-255.
- (15) Rye, H. S.; Yue, S.; Wemmer, D. E.; Quesada, M. A.; Haugland, R. P.; Mathies, R. A.; Glazer, A. N. *Nucl. Acids Res.* **1992**, *20*, 2803-2812.

- (16) Yue, S. T.; Johnson, I. D.; Huang, Z.; Haugland, R. P. *Unsymmetrical cyanine dyes with a cationic side chain*; Molecular Probes, Inc.: U.S. Patent 5,321,130, 1994.
- (17) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561.
- (18) Staerk, D.; Hamed, A. A.; Pedersen, E. B.; Jacobsen, J. P. *Bioconj. Chem.* **1997**, *8*, 869-877.
- (19) Jacobsen, J. P.; Pedersen, J. B.; Hansen, L. F.; Wemmer, D. E. *Nucl. Acids Res.* **1995**, *23*, 753-760.
- (20) deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909-7916.
- (21) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534-3535.
- (22) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471.
- (23) Förster, T.; Sinanoglu, O., Ed.; Academic: New York, 1965, pp 93-137.
- (24) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; Freeman: San Francisco, 1980; Vol. 2.
- (25) Stryer, L.; Haugland, R. P. *Proc. Natl. Acad. Sci USA* **1967**, *58*, 719-730.
- (26) Mergny, J. L.; Bourtine, A. S.; Garestier, T.; Belloc, F.; Rougee, M.; Bulychov, N. V.; Koshkin, A. A.; Bourson, J.; Lebedev, A. V.; Valeur, B.; Thuong, N. T.; Helene, C. *Nucl. Acids Res.* **1994**, *22*, 920-928.
- (27) Sixou, S.; Szoka, F. C.; Green, G. A.; Giusti, B.; Zon, G.; Chin, D. J. *Nucl. Acids Res.* **1994**, *22*, 662-668.
- (28) Tyagi, S.; Kramer, F. R. *Nat. Biotech.* **1996**, *14*, 303-308.
- (29) Tyagi, S.; Bratu, D. P.; Kramer, F. R. *Nat. Biotech.* **1998**, *16*, 49-53.
- (30) Kostrikis, L. G.; Tyagi, S.; Mhlana, M. M.; Ho, D. D.; Kramer, F. R. *Science* **1998**, *279*, 1228-1229.

- (31) Zlokarnik, G.; Negulescu, P. A.; Knapp, T. E.; Mere, L.; Burres, N.; Feng, L. X.; Whitney, M.; Roemer, K.; Tsien, R. Y. *Science* **1998**, 279, 84-88.
- (32) Ebata, K.; Masuko, M.; Ohtani, H.; Kashiwasakejibu, M. *Photochem. Photobiol.* **1995**, 62, 836-839.
- (33) Robertson, J. M.; White, J. G. *J. Chem. Soc.* **1947**, 358-368.
- (34) Cho, N.; Asher, S. A. *J. Am. Chem. Soc.* **1993**, 115, 6349-6356.
- (35) Chen, F. M. *Nucl. Acids Res.* **1983**, 11, 7231-7250.

Chapter 3:

Polyamide Conjugates for Affinity Purification of DNA

3.1 Background

An important step in many molecular biology manipulations is the purification of double stranded DNA. A number of diverse purification schemes have been developed, particularly for plasmid DNA. Sequence-specific purification and screening techniques are also being developed. The protein RecA, for instance, has been used to search for rare DNA sequences,¹ and for the purification of DNA fragments containing these sequences.² A second technique of particular interest is triple helix mediated affinity capture.^{3,4} In this technique, a triple helix forming oligonucleotide is designed to bind to a plasmid or other DNA fragment. The triple helix forming oligonucleotide is used as a tether to capture the triple helical complex using a solid phase resin or magnetic bead technology. For some applications, this approach is limited by the sequence recognition repertoire available to the triple helix approach.^{5,6} Given the more general rules of recognition and the high affinity and specificity of minor groove binding polyamides,^{7,8} we sought to test whether polyamides were competent for sequence specific affinity capture of double stranded DNA (dsDNA). Like the triple helix approach, use of minor groove binding polyamides shares the advantage of capturing double stranded DNA ready for use in further manipulations.

Successful design of such a system of polyamide affinity capture reagents as with their triple helical counterparts could provide tools for the purification of plasmid to cosmid DNA, or for the isolation of genes in cDNA libraries, bacterial artificial chromosome

(BAC) libraries, or in crude genomic samples.^{9,10} One could envision utilization of such a system for the cloning of homologous genes in concert with degenerate PCR methodologies. These technologies compare favorably with current genomic screening approaches that usually rely on *in situ* filter hybridization.¹¹

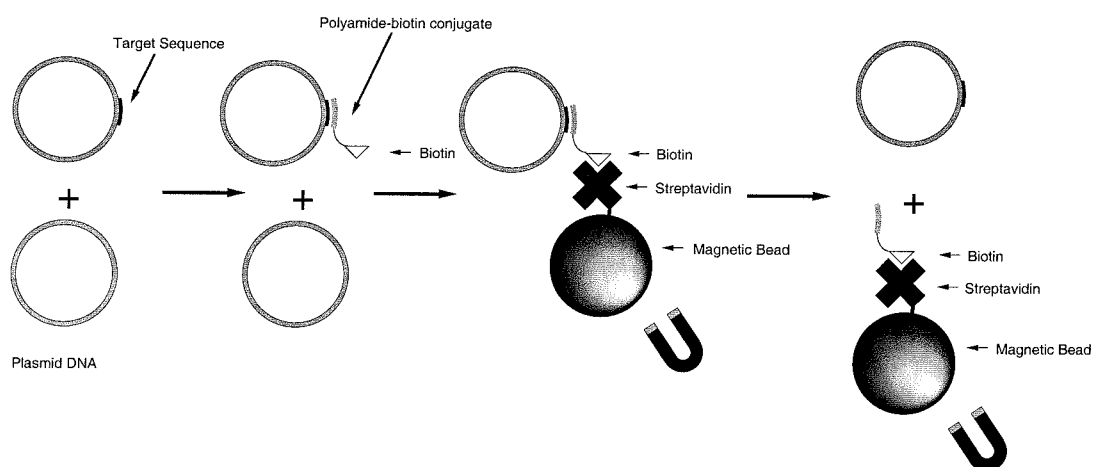


Figure 1. Polyamide affinity capture. DNA fragments containing a polyamide binding site are separated from remaining DNA through interaction with magnetic beads recognizing an affinity tail on the compound, in this case a biotin-streptavidin interaction.

Several approaches have been used in the literature for the purification of DNA by the triple helix affinity capture methodology. The first involves tethering a biotin moiety to the third strand oligonucleotide, and capturing the triple helix with streptavidin immobilized on either resin or magnetic beads.^{3,4} The second approach, using a strand displacement approach rather than third strand, involved the conjugation of a His₆ tag to a polyamide nucleic acid (PNA) strand, and isolation using a nickel binding column.¹² We designed polyamide conjugates based upon both of these approaches. A polyamide of the sequence ImPyPyPy- γ -ImPyPyPy- β -Dp was chosen as a parent structure for these studies, as it had

been shown to bind DNA with high affinity, and good specificity.⁷ A series of conjugates were prepared with biotin or the His₆ tag that differed in the linker length. These conjugates were assayed for the ability to capture a moderately sized restriction fragment (~250bp) containing either the targeted match binding site, or a single base mismatch. The length of the polyamide binding site must correlate with the length of the fragment to be captured. Figure 3 shows the number of base pairs necessary to specify a unique site in a given length of DNA.¹³ On the restriction fragment level, this would correspond to a polyamide binding to a 4-5 base pair binding site. Accounting for the pyrrole-pyrrole degeneracy in the compounds used, this value must be increased. The polyamide used binds to the sequence 5'-WGWWCW-3'. This compound recognizes 10 out of the 2080 possible six base pair binding sites, effectively reducing the length of the target DNA to approximately 200 bp.

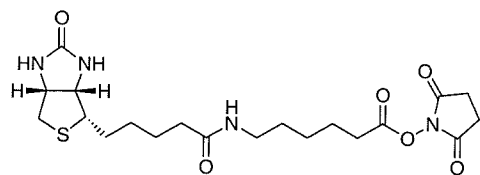
Target DNA	genome length	bp of unique site
Plasmids	2-5x10 ³	6-7
Bacteriophage λ	4.86x10 ⁴	8-9
Bacteria (<i>E. coli</i>)	4.70x10 ⁶	11-12
Yeast (<i>S. cerevisiae</i>)	1.35x10 ⁷	13
Fly <i>D. melanogaster</i>	1.65x10 ⁸	14-15
Human Genome	2.9x10 ⁹	16-17

Figure 2. Base pairs necessary to recognize unique sites in various experimentally relevant organisms.

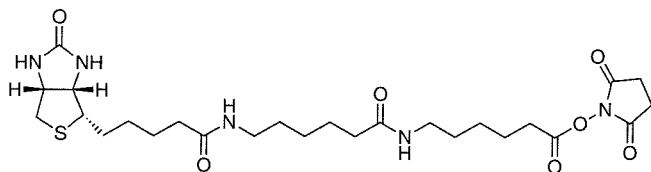
3.2 Synthesis of Polyamide Conjugates

Polyamide conjugates were prepared with biotin or metal-binding His₆ tags by standard solid phase synthetic methodologies. The biotin conjugates were prepared by cleavage of the polyamide from the solid support using various diamines. The resulting

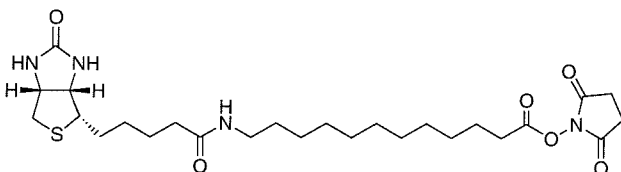
primary amines were then coupled to biotinylating reagents, either commercially available, or prepared as shown in figure 4.



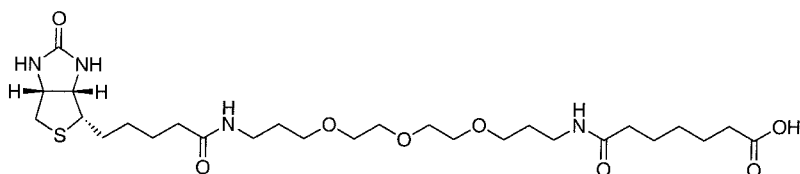
1. BIOTIN-X-NHS ESTER



2. BIOTIN-XX-NHS ESTER



3. BIOTIN-C11-NHS ESTER



4. BIOTIN-21-COOH

Figure 3. Biotinylating reagents. 1, 2, and 3 are commercially available. 4 was synthesized as described below.

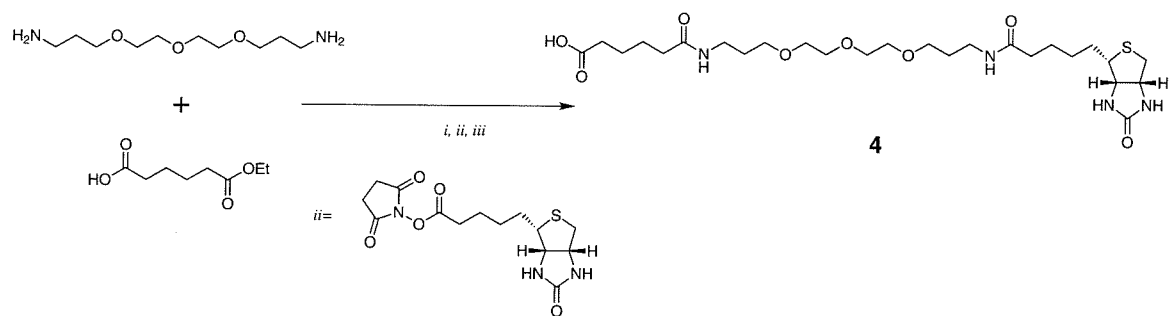


Figure 4. Synthesis of biotin linker derivative 4. *i*) dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBT); *ii*) biotin-NHS ester; *iii*) NaOH/MeOH.

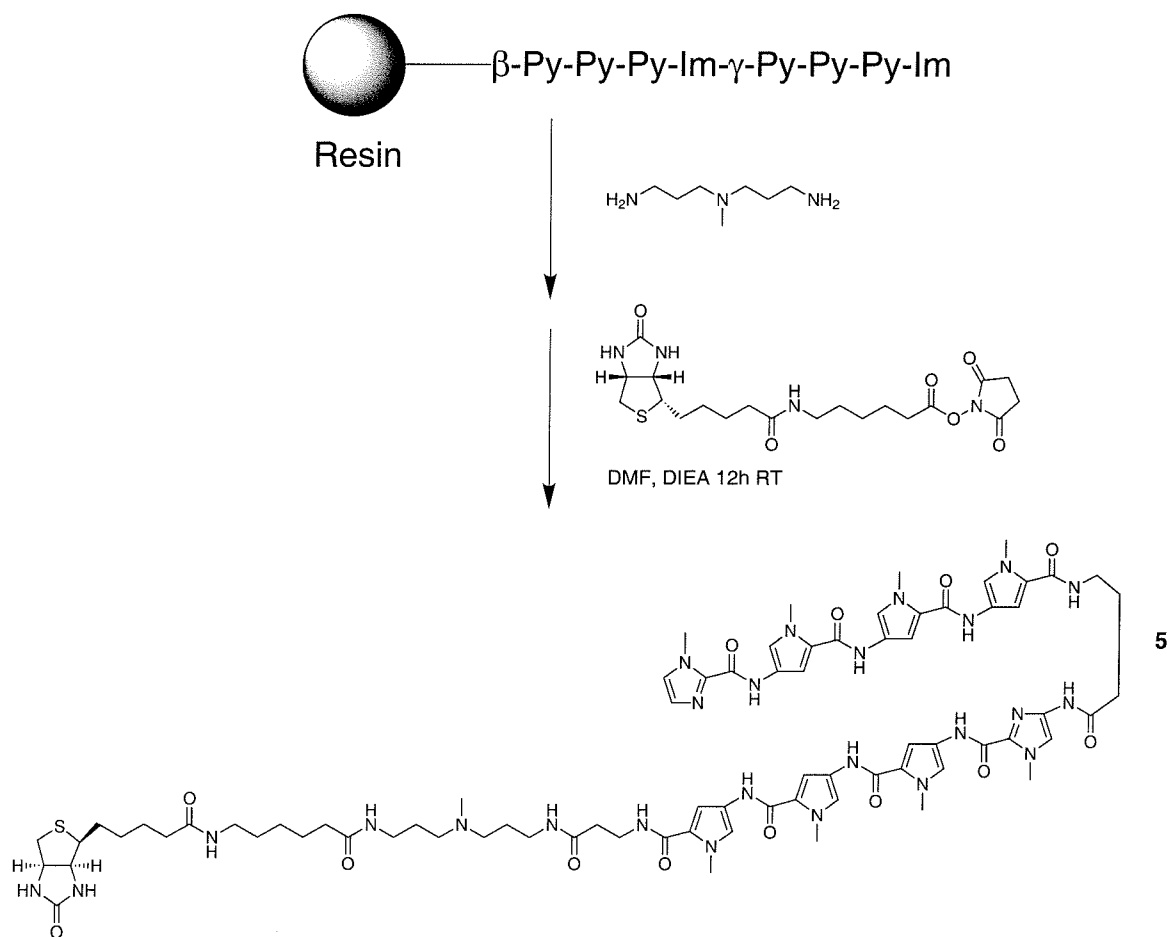


Figure 5. Synthesis of Biotin Conjugates. Polyamide resin was cleaved with N-methyl-N-dipropylamine and HPLC purified. The purified polyamide treated with N-hydroxysuccinimide ester of biotin-hexanoic acid, and the conjugate purified by HPLC.

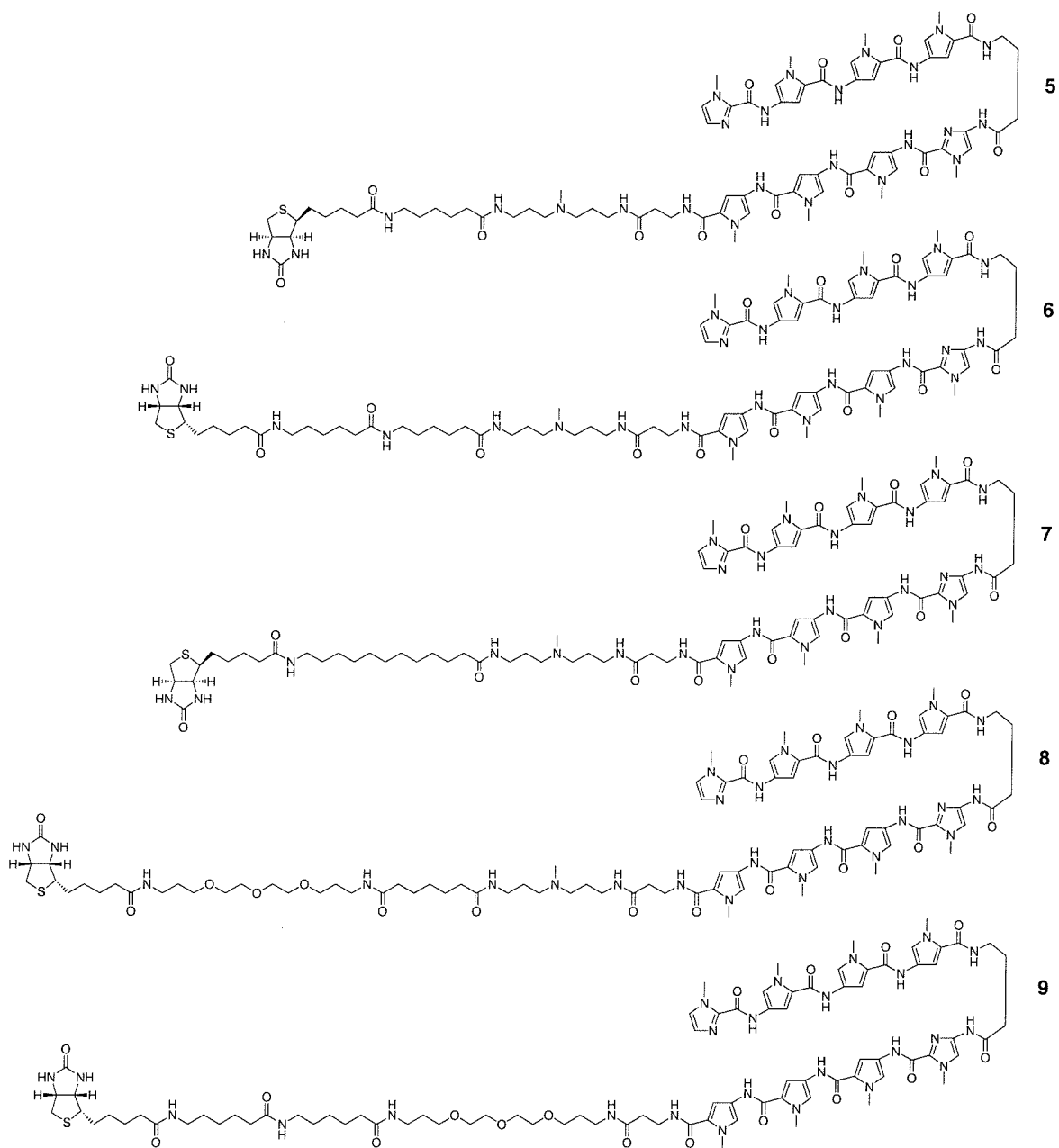


Figure 6. ImPy3-b-ImPy3-b biotin conjugates used in this study. 5.

ImPyPyPy- γ -ImPyPyPy- β -Dp-X-biotin; **6.** ImPyPyPy- γ -ImPyPyPy- β -Dp-XX-biotin;

7. ImPyPyPy- γ -ImPyPyPy- β -Dp-11-biotin **8.** ImPyPyPy- γ -ImPyPyPy- β -Dp-21-biotin;

9. ImPyPyPy- γ -ImPyPyPy- β -15-XX-biotin.

The histidine conjugate was prepared by solid phase synthesis using commercially available Boc-His-DNP on PAM-Glycine resin. The His₆ resin was extended with 3

linker amino acid residues, followed by normal polyamide synthesis. The resulting compound was cleaved from the resin using HF, and precipitated with ethyl ether. The resulting conjugate was purified by reversed phase HPLC.

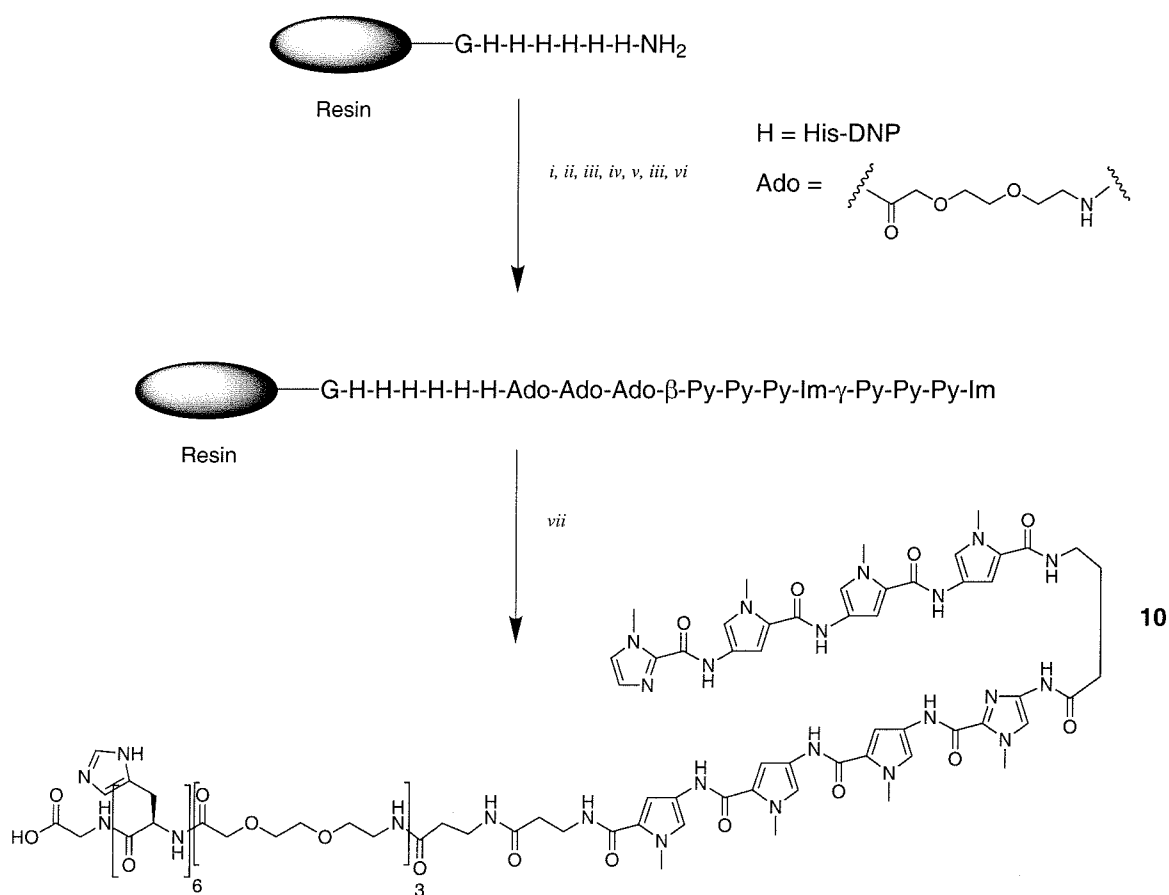


Figure 7. Synthesis of His₆ conjugate. DNP-protected His₆Gly PAM resin treated as follows: *i*) a. Fmoc-Ado-COOH, HBTU, DMF, DIEA b. 20% piperidine; 3X; *ii*) a. Boc-b-Ala-COOH, HBTU, DMF, DIEA b. TFA; *iii*) a. Boc-Py-COOBt b. TFA; *iv*) Boc-Im-COOH, HBTU, DIEA, DMF b. TFA; *v*) a. Boc-γ-COOH, HBTU b. TFA; *iv*) Im-COOH, HBTU, DMF, DIEA; *vii*) HF cleavage, Et₂O precipitation.

3.3 Results and Discussion

For convenience of assay, compounds were tested in parallel experiments for their ability to capture an *EcoRI/PvuII* restriction fragment derived from either pJT8 or pUC19. The broad strategy of experiments is outlined in figure 8 for the biotin compounds, but the

same strategy was used for the His₆ tagged compounds. The fragments were chosen such that the pUC19 sequence did not contain a 5'-WGWWCW-3' match site for the polyamide, but only single base pair mismatches.

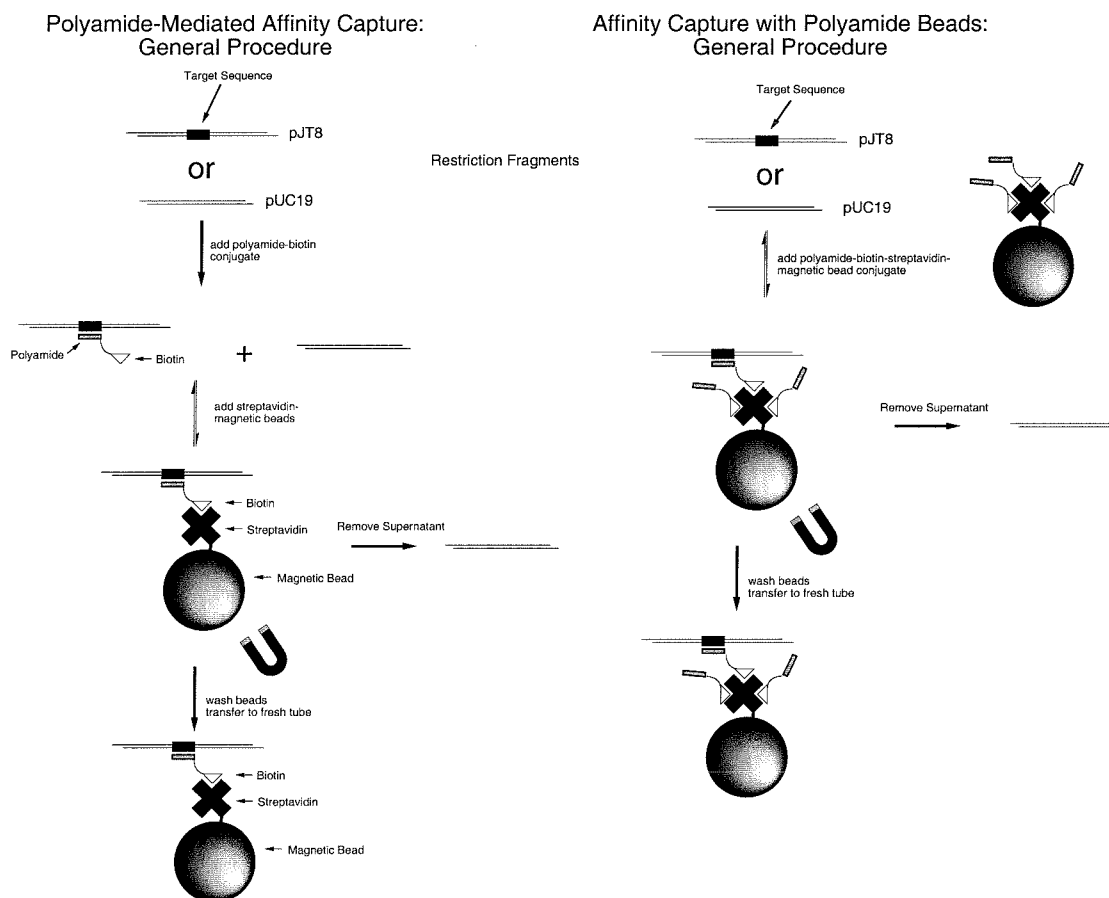


Figure 8. Experimental capture procedures. In tandem, 3-10,000 cpm of radiolabeled *EcoRI/PvuII* restriction fragment from either pUC19 or pJT8 was equilibrated with either polyamide-biotin conjugate or preformed polyamide-biotin-streptavidin magnetic beads. The beads were separated with a magnetic stand and the supernatant removed. The captured beads were washed twice with the equilibration buffer then transferred to a new tube before scintillation counting.

pUC19 EcoRI/PvuII (234 bp)

5' -AATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAA
 3' -TTAAGCTCGAGCCATGGGCCCCTAGGAGATCTCAGGTGGACGTCCGTACGTT

GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCAC
 CGAACCGCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGAG

AATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAA
 TTAAGGTGTGTTGTATGCTCGGCCTTCGTATTTACATTTTCGGACCCACGGATT

TGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTTTCAGTCGG
 ACTCACTCGATTGAGTGTAATTAACGCAACGCGAGTGACGGGCGAAAGGACAGCC

GAAACCTGTCGTGCCAG-3'
 CTTTGGACAGCACGGTC-5'

pJT8 EcoRI/PvuII (302 bp)

5' -AATTCGAGCTCGGTACCCGGTTAGTATTTGGATGGGCCTGGTT**AGTACTTG**
 3' -TTAAGCTCGAGCCATGGGCCAATCATAAACCTACCCGACCAAT**CATGAAC**

GATGGGAGACCGCCTGGGAATACCAGGTGTCGTATCTTAACTTAAGGCGTAATCAT
 CTACCCTCTGGCGGACCCCTTATGGTCCACAGCATAGAATTGAATTCGCATTAGTA

GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCACACAACATA
 CCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGAGTTAAGGTGTGTTGTAT

CGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCAC
 GCTCGGCCTTCGTATTTACATTTTCGGACCCACGGATTACTCACTCGATTGAGTG

ATTAATTGCGTTGCGCTCACTGCCCCGCTTTCAGTCGGGAAACCTGTCGTGCCAG-3'
 TAATTAACGCAACGCGAGTGACGGGCGAAAGGACAGCCCTTTGGACAGCACGGTC-5'

Figure 9. Sequences of EcoRI/PvuII restriction fragments of pJT8 and pUC19.

Overall, the biotinylated compounds showed an ability to capture DNA, but with little specificity, even at low concentrations. Depending on the conditions of the experiment, 20-90% of the DNA could be captured with streptavidin coated magnetic beads. Results were quite similar for each of the compounds tested, so results are shown for conjugate **6**, ImPyPyPyPy- γ -ImPyPyPy- β -Dp-XX-biotin. Figure 10 shows the result of a single capture experiment with polyamide 6.

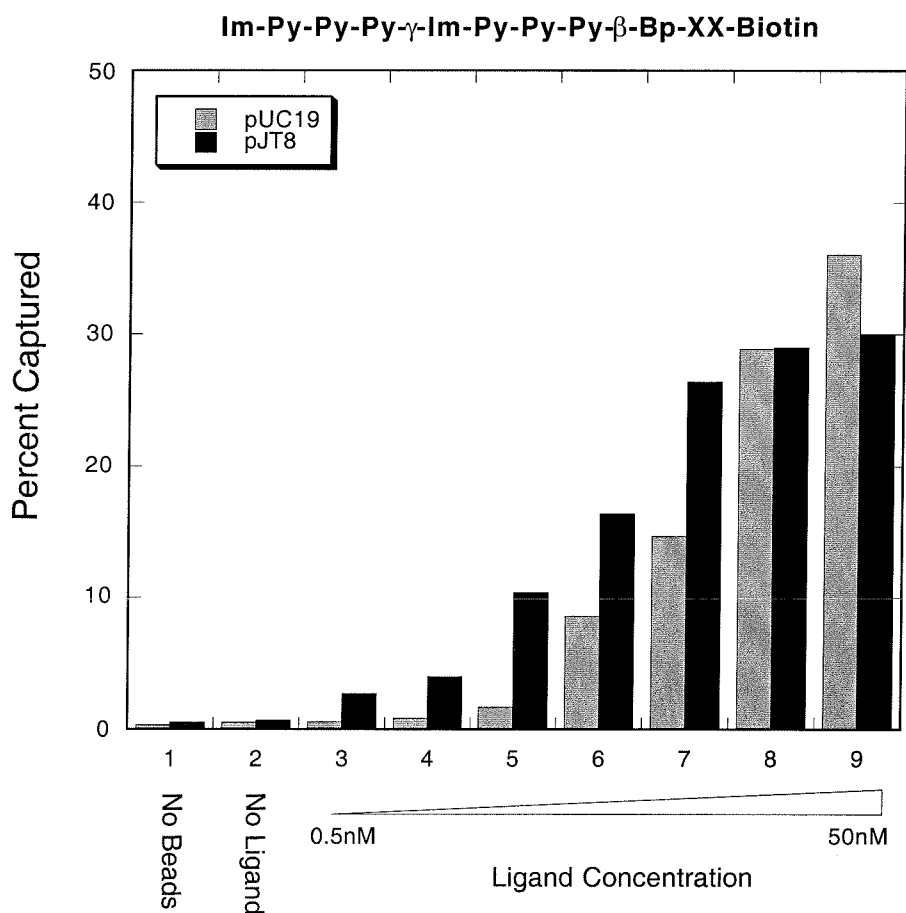


Figure 10. Representative capture data. Radiolabeled restriction fragments were treated with biotinylated polyamide 6, and equilibrated for 12 hours under standard conditions. The mixture was treated with 10 μ L of streptavidin coated magnetic beads and separated magnetically, followed by two washes with 1X buffer. Experiments were performed in tandem.

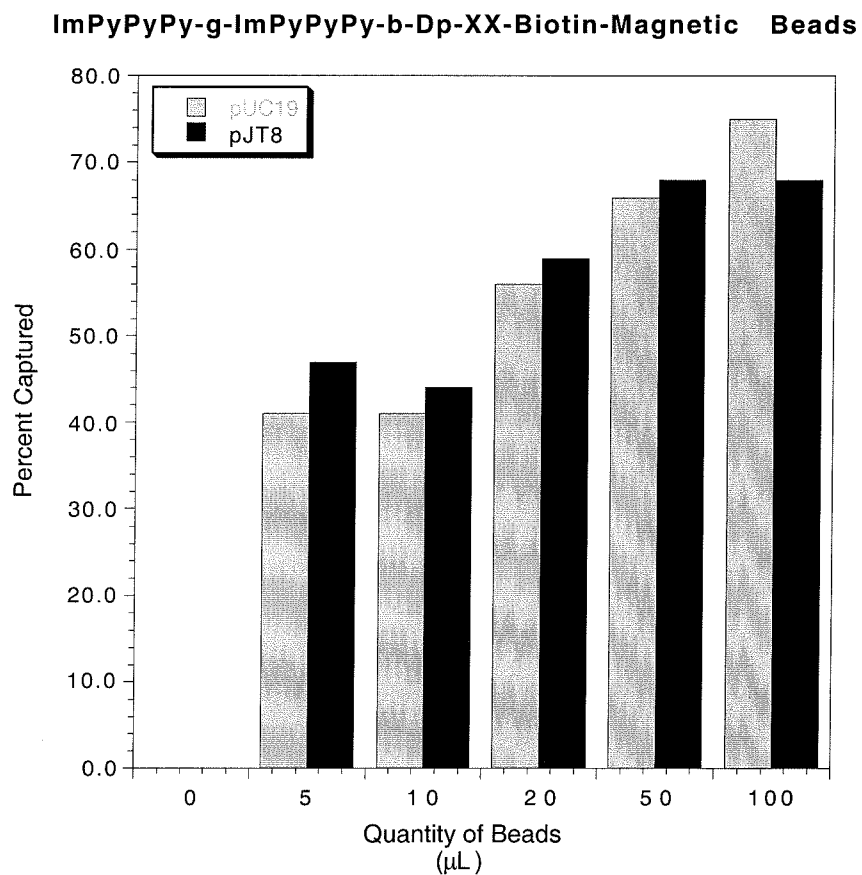


Figure 11. Capture experiment using polyamide beads. Preformed polyamide-XX-biotin-streptavidin-beads were added to solutions of restriction fragments 5,000 cpm, incubated for 12 hours, and separated magnetically. The beads were washed twice with 1X TKMC and counted.

The success of the preformed triplex affinity capture approach used by Smith and coworkers led us to try capture experiments with preformed polyamide conjugates.³ As these authors observed, increased quantities of beads utilized in the experiment led to

more efficient levels of capture even though the smallest amount of bead should have the capacity to bind all of the radiolabeled DNA.

Given the decrease in observed specificity of these conjugates, we determined the affinity binding constants for a biotinylated derivative for DNA. For compound **6**, the affinity we measured was within a factor of 10 of the parent hairpin compound. The DNase I footprinting titration is shown in figure 13. The same footprinting experiment was then performed in the presence of free streptavidin protein, shown in figure 14. In the presence of streptavidin, the compound binds with a drastic reduction in specificity. Streptavidin is a tetrameric protein. We reasoned that the decrease in binding specificity could result from multidentate binding interactions. Computer modeling of the streptavidin tetramer showed that it was clearly possible to bind a large piece of DNA multiple times with a tetrameric streptavidin polyamide complex. Indeed, the DNase I footprinting showed drastically reduced specificity in the presence of streptavidin. Interestingly, the enhanced cleavage sites, and periodic cleave pattern of the streptavidin gel, suggest that the polyamide-biotin-streptavidin complex might be forming a histone-like bent DNA structure. This could be an alternative mechanism for generating bent DNA structures, but was not further explored.

With a working model, that multiple binding events had lowered the apparent specificity, an experiment was designed to make “monomeric” streptavidin through the titration of free biotin. Unfortunately, this experiment did not improve the selectivity, most likely since the free biotin binds in a random fashion to the streptavidin, giving a statistical mixture of bound complexes. A second possibility for the decrease in specificity is that the streptavidin-polyamide complex facilitates interactions between the DNA and the streptavidin protein. Although free streptavidin has no discernible binding to the DNA at the concentrations used in the DNase experiment, it is possible that weak affinity could be observed when the DNA is proximally tethered.

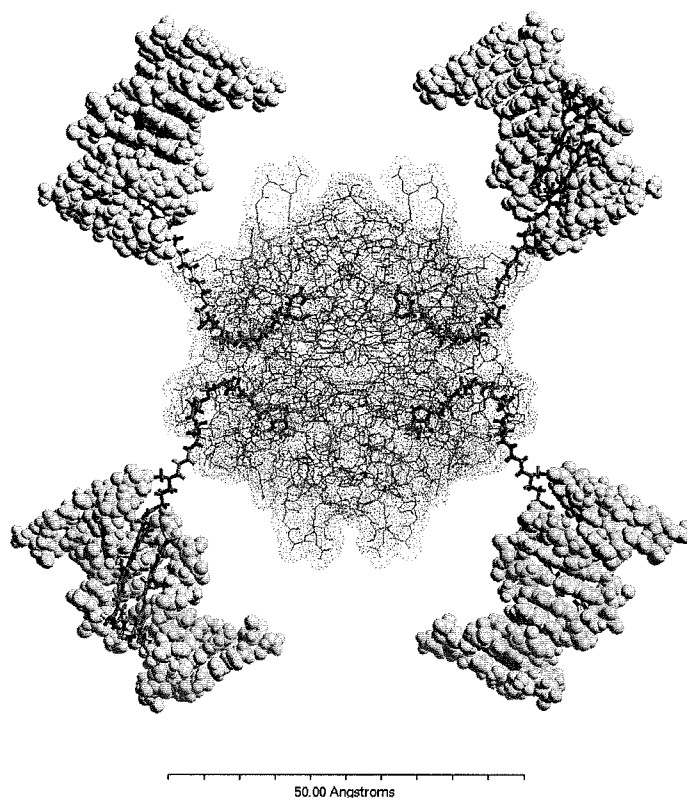


Figure 12. Model of a streptavidin-biotin-polyamide complex bound to DNA. Figure was generated using Biograf (MSI, Carlsbad, CA) from coordinates of a streptavidin-biotin tetrameric complex.¹⁴ The hairpin ImPyPy- γ -PyPyPy- β -Dp-X-biotin polyamide used for the model originated from an NMR structure from Wemmer and coworkers¹⁵ that was edited to give the γ turn, the β tail, and the biotin hexanoic acid tail.

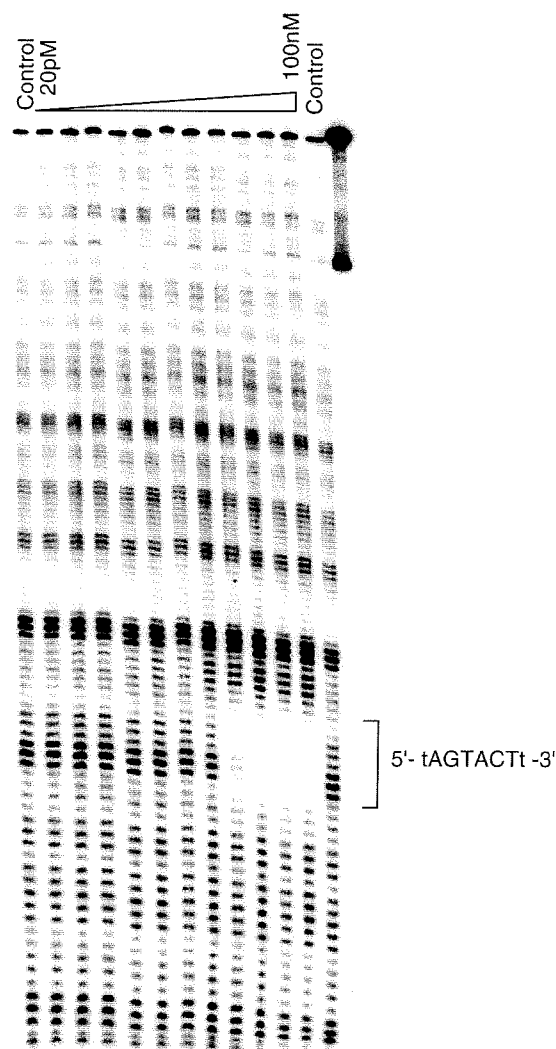
ImPyPyPy- γ -ImPyPyPy- β -Bp-XX-Biotin

Figure 13. DNase Footprint titration of ImPyPyPy- γ -ImPyPyPy- β -Dp-XX-Biotin. 32 P-labeled pJT8 EcoRI/PvuII restriction fragment. Lanes 2-13 contain increasing polyamide with 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, and 100nM concentrations respectively. Reactions were carried out in 40 μ L total volumes with final solution conditions of 25nM streptavidin, 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, and pH 7.0. Polyamide was equilibrated for 18 hours prior to reaction with DNase I. Lanes marked control contain no polyamide. Streptavidin lane contains 25nM streptavidin. "A" reaction run on same gel is not shown.¹⁶

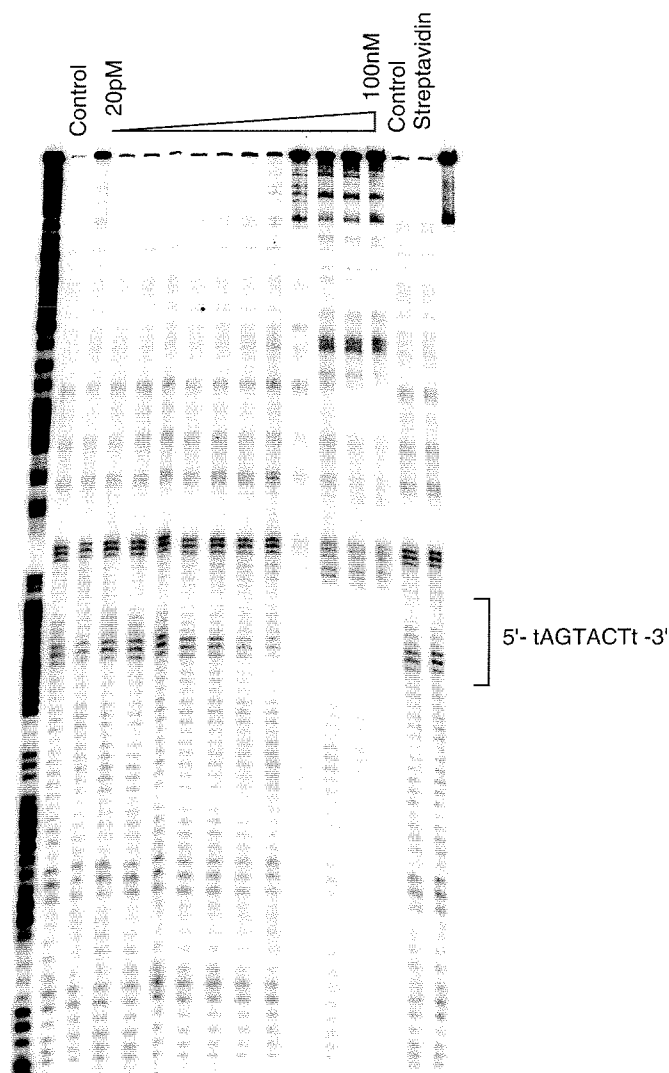
ImPyPyPy- γ -ImPyPyPy- β -Bp-XX-Biotin in Presence of Streptavidin

Figure 14. DNase footprinting titration of ImPyPyPy- γ -ImPyPyPy- β -Dp-XX-Biotin in the presence of streptavidin. 3' 32 P-labeled pJT8 EcoRI/PvuII restriction fragment. Lanes 3-14 contain increasing polyamide with 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, and 100nM concentrations respectively. Reactions were carried out in 40 μ L total volumes with final solution conditions of 25nM streptavidin, 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, and pH 7.0. Polyamide was equilibrated for 18 hours prior to reaction with DNase I. Lanes marked control contain no polyamide or streptavidin. Streptavidin lane contains 25nM streptavidin. Lane 1 contains an "A" reaction.¹⁶

The second means attempted to overcome this multidentate interaction with DNA was to change the nature of the capture technique. To this end, a single His₆ tagged

polyamide was synthesized. The conjugate was designed with a lengthy hydrophilic linker to avoid interactions with the beads. Experimentally, this became more feasible as Ni(II) bound magnetic agarose beads became commercially available in mid-1998 from Qiagen, Inc. This polyamide was able to capture 85% of a radiolabeled restriction fragment containing either a match or mismatch, with very modest specificity using either assay format.

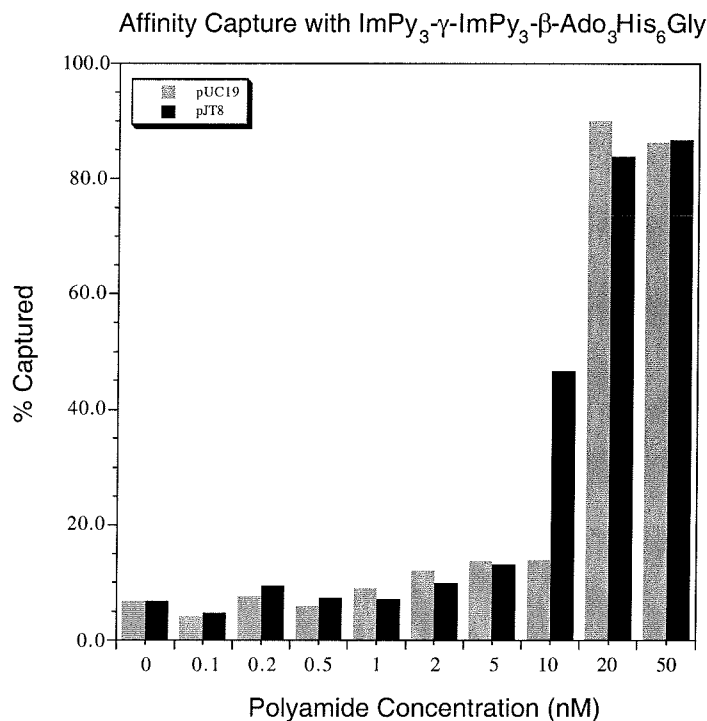


Figure 15. Polyamide mediated affinity capture with ImPyPyPy-γ-ImPyPyPy-β-Ado₃His₆Gly. Polyamides were incubated with either pJT8 or pUC19 (≈3,000 cpm) for 12 hours at room temperature. 10 μL of Ni-NTA magnetic beads (suspension from Qiagen) were added to give a final volume of 400 μL. The mixture was mixed by inversion for 30 minutes, followed by magnetic separation. The beads were washed twice with 1X TKMC buffer and the residual radioactivity determined by scintillation counting.

Unlike the biotin experiments, the His₆ system had the inherent advantage that the polyamide-DNA complex could be eluted from the beads with high concentrations of imidazole. Figure 15, for instance, shows the capture of the restriction fragments for pJT8 and pUC19 with preformed polyamide beads. Preformed beads showed little selectivity in the capture of the the match fragment (pJT8) over the mismatch fragment (pUC19). At best

a three-fold specificity could be achieved, but the error in the experiment reduces the significance of this observation.

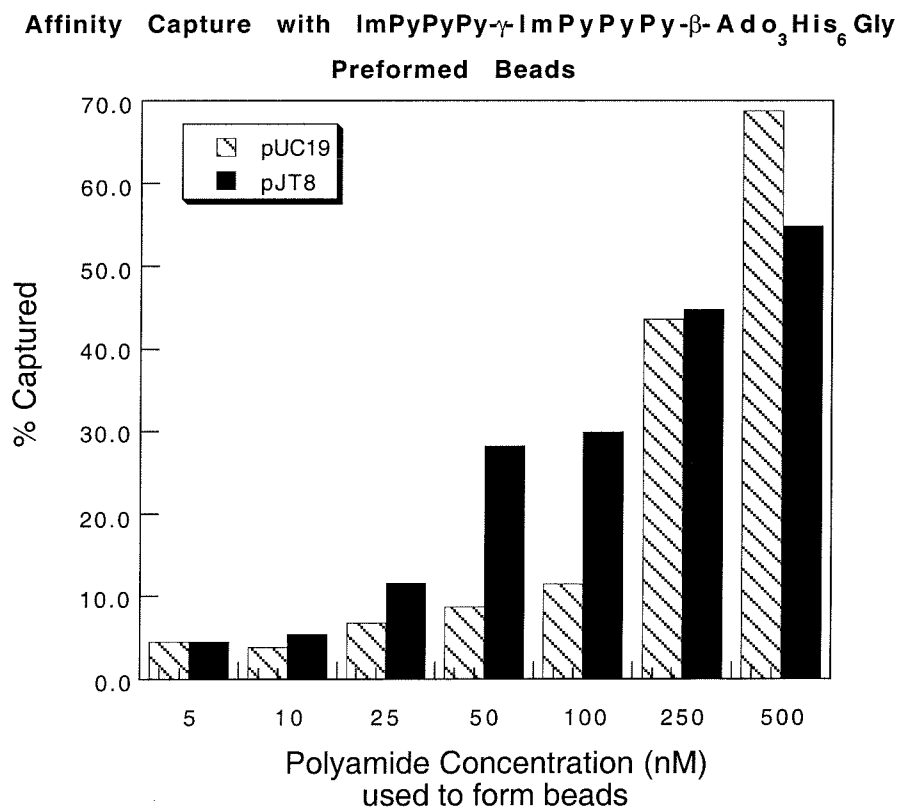


Figure 16. Polyamide mediated affinity capture with preformed beads of ImPyPyPy- γ -ImPyPyPy- β -Ado₃His₆Gly. Polyamide beads were preformed by adding 10 μ L of Ni-NTA agarose bead suspension to 200 μ L of the appropriate concentration of polyamide-His₆ conjugate. The tubes were mixed for 30 minutes and separated magnetically. The beads were washed with water (2x200 μ L), suspended in 20 μ L of water and added to tubes containing either pJT8 or pUC19 (\approx 3,000 cpm) to give a final volume of 400 μ L. Final buffer conditions were, 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, and pH 7.0. The tubes were mixed for 30 minutes and separated magnetically. The beads were washed twice with 1X TKMC buffer and the remaining radioactivity determined by scintillation counting.

A significant advantage of the His₆ system over the biotin system is the ability to remove the polyamide-DNA complex from the magnetic beads under mild non-denaturing

conditions. This could also be possible in the biotin system through the introduction of a cleavable linker, or through the development of a system for switching the polyamide binding affinity. For the His₆ system, a polyamide-DNA complex could be efficiently eluted from a bead using either high concentrations of free imidazole (200mM) or by decreasing the pH below 4. Figure 16 shows the capture, release and recovery of the *EcoRI/PvuII* restriction fragments derived from pJT8 and pUC19. Polyamides could capture and release these restriction fragments in overall yields of approximately 70%, again with little specificity for either restriction fragment.

To confirm that addition of the His₆ tag had not disrupted the polyamide's affinity or specificity, the compound's affinity was determined using quantitative DNase I footprinting. Footprinting on the pJT8 restriction fragment reveals that the compound still binds with high affinity, although some 40-fold reduced from the parent hairpin. The specificity of the conjugate compares quite favorably to that of the parent hairpin for both the biotinylated and His₆ labeled compounds.

Table 1. Apparent equilibrium association constants (M⁻¹) of affinity capture conjugates.^a

Polyamide Conjugate	5'-ttAGTATTTtg-3'	5'-ttAGTACTtg-3'
ImPyPyPy-γ-ImPyPyPy-β-Dp ¹	3.7 x 10 ¹⁰	4.1 x 10 ⁸
ImPyPyPy-γ-ImPyPyPy-β-Dp-XX-Biotin (6)	1.7 x 10 ⁸	< 5 x 10 ⁶
ImPyPyPy-γ-ImPyPyPy-β-Dp-XX-Biotin•Streptavidin (6)	1.5 x 10 ⁸	7.8 x 10 ⁷
ImPyPyPy-γ-ImPyPyPy-β-Ado ₃ His ₆ Gly (10)	8.3 x 10 ⁸	1.3 x 10 ⁸

^a Quantitative DNase I footprinting reactions were carried out on a 3'-labeled *EcoRI/PvuII* restriction fragment derived from plasmid pJT8 as described.¹⁷ Data represent single trials.

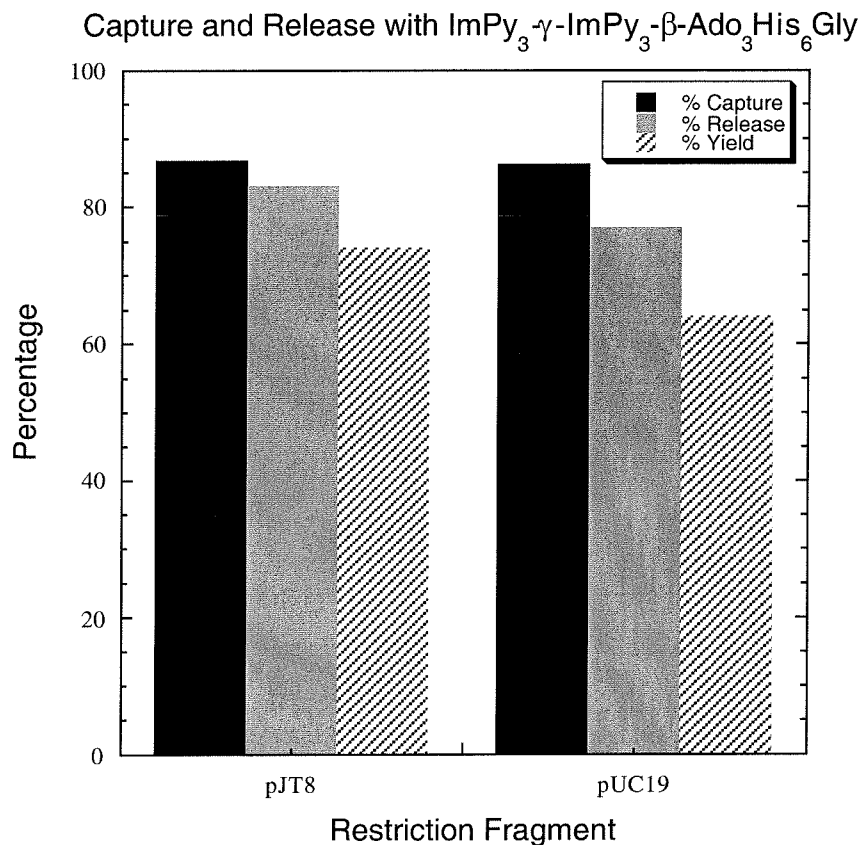


Figure 17. Capture and release of labeled restriction fragments.

Polyamide beads were formed by mixing 10 μL of agarose Ni-NTA beads with 200 μL of 1 μM $\text{ImPyPyPy}\text{-}\gamma\text{-ImPyPyPy}\text{-}\beta\text{-Ado}_3\text{His}_6\text{Gly}$ and inverting for 30 minutes. The beads were washed twice with water, and suspended in 400mL of 1X TKMC buffer containing either 5,000 cpm of either the pJT8 or the pUC19 *EcoRI/PvuII* end-labeled restriction fragment. After mixing for 30 minutes, the beads were removed by magnetic separation and washed twice with 1X TKMC. The washed beads were counted to give the % capture. The DNA-polyamide complex was eluted from the beads with 200 μL of 200 μM imidazole in 5X TKMC buffer in two 15 minute washes. The washes were combined and counted to give the % release. The percent yield is calculated as the product of the percent capture and the percent release.

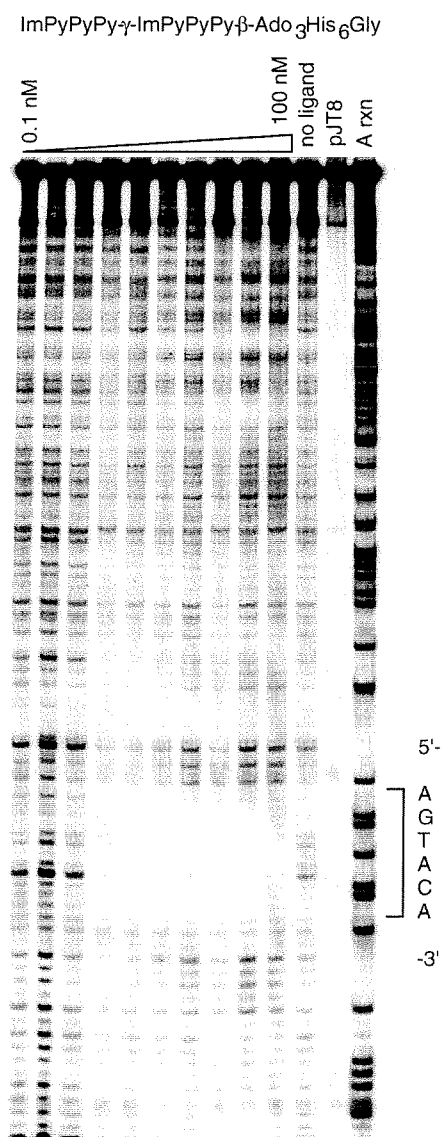


Figure 18. Quantitative DNase I footprinting of ImPy₃- γ -ImPy₃- β -Ado₃His₆Gly. 3' ³²P-labeled pJT8 EcoRI/PvuII restriction fragment. Lanes 1-10 contain increasing polyamide with 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, and 100nM concentrations respectively. Reactions were carried out in 400mL total volumes with final solution conditions of 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, pH 7.0. Polyamide was equilibrated for 18 hours prior to reaction with DNase I.

3.4 Conclusions

This work has shown that polyamides can efficiently capture DNA from solution. The histidine tagged polyamide had the added benefit of convenient release of the polyamide-DNA complex from the magnetic beads. While the sequence specificities found in these experiments were disappointing, the general approach is valid. Polyamides containing hydroxypyrrole to break the A-T degeneracy in the binding site should improve the selectivity. Increasing the length of polyamide and its cognate binding site should also improve the selectivity of the system. In comparison to the triple helix approach, polyamide mediated affinity capture is in its infancy. The technique should improve as binding site size and specificity of polyamides are increased. Even without increased sequence specificity, these conjugates could find usage for the separation of dsDNA from ssDNA in PCR applications. More specific conjugates in the future could be important reagents for biochemical purification of DNA.

3.5 Experimental

3.5.1 Materials

Streptavidin-coated magnetic beads were obtained from either Dynal or from Promega. The Dynal (M-280) beads gave superior results to those obtained from Promega. Agarose Ni-NTA magnetic beads were obtained from Qiagen.

Restriction Fragments:

Restriction fragments were excised from pJT8 or pUC19 using the combination of restriction endonucleases, *EcoRI* and *PvuII*. The fragments were 3' end-labeled using Sequenase polymerase and α -³²P-dATP and α -³²P-TTP. The restriction fragments were purified on a 7% non-denaturing polyacrylamide gel, visualized by autoradiography, cut

from the gel, and eluted by soaking overnight in 1X tris-EDTA (TE) buffer. The purified fragments were ethanol precipitated, washed, and aliquoted for experiments.

TKMC Buffer: Buffer was prepared as a 5X stock solution containing: 50 mM Tris-HCl, 50 mM KCl, 50 mM MgCl₂, 25 mM CaCl₂, at pH 7.0. Capture and DNase I footprinting reactions were carried out with 1X TKMC containing 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, 5 mM CaCl₂, at pH 7.0.

3.5.2 General Procedures

Polyamide titrations were set up in 600 μ L eppendorf tubes. Tandem sets of experiments were performed with either the fragment from pJT8 or the fragment from pUC19. The samples were allowed to equilibrate from 1 hour to 24 hours, then magnetic beads were added for 15 minutes to 12 hours. The beads were separated from the supernatant using a magnet, and the supernatant removed with a 200 μ L pipet. The beads were washed with 1X or 5X buffer twice. All samples were saved, and the distribution of the radiolabeled fragment determined by scintillation counting. The Dynal magnetic beads suppressed signals from radioactive samples to the extent of approximately 15%, fairly uniformly.

Polyamide-beads: In other experiments, preformed polyamide-bead conjugates were prepared by soaking magnetic beads in a 1 μ M solution of the polyamide, followed by two washes with the appropriate reaction buffer.

Polyamide Synthesis: The parent hairpin was synthesized on β -PAM resin to give Resin- β -PyPyPyIm- γ -PyPyPyIm using standard protocols¹⁷ as described with the exception that deprotections were carried out with neat trifluoroacetic acid and dichloromethane washes were omitted.

Resin cleavage: The resin was cleaved with either of two diamines, *N*-methyl-*N,N*-dipropylamine or 4,7,10-Tioxa-1,13-tridecanamine. The cleaved products were purified by HPLC using a C18 reverse phase column and a gradient of 0.25% acetonitrile per minute in the presence of 0.1% TFA. Pure fractions determined by analytical HPLC were lyophilized and aliquoted for coupling reactions.

Biotinylation: 1-10 mg of each polyamide was dissolved in dimethylformamide (DMF), 100 μ L and *N,N*-diisopropylethylamine (DIEA), 25 μ L. Biotin-NHS esters were added as a solution in 50 mL DMF. Coupling was allowed to proceed for 12 hours at RT with occasional stirring. The coupling reactions were monitored by analytical HPLC. Reactions with commercially available NHS esters coupled with greater than 90% efficiency. The crude reaction mixtures were purified using the same HPLC conditions to yield conjugates 5, 6, 7, and 9. The conjugates were characterized by mass spectrometry, and in some cases by ¹H-NMR spectroscopy. Mass spectrometry results are summarized in the following table:

Polyamide conj.	Calc'd Mass (M+H ⁺)	Exp't. (M+H ⁺)
5	1604.8	1604.9
6	1717.9	1718.6
7	1688.9	1687.7
8	1834.9	1835.0
9	1792.9	1793.1
10	2524.1	2524.5

Biotin-21-COOH (**24**): 4.36g of adipic acid monoethyl ester (25 mmol) was activated with 3.38g of hydroxybenzotriazole (HOBt) and 5.16g of dicyclohexylcarbodiimide (DCC) in 20 mL of DMF for 5 hours. This cloudy mixture was filtered into a stirred solution of 4,7,10-Trioxa-1,13-tridecaneamine 22.03g (100mmol, 4 eq.) in 50 mL DMF. The solution was stirred overnight. The reaction was diluted with ethyl acetate 250 mL and partitioned against brine. The ethyl acetate was washed with dilute HCl (pH 3), then dried and evaporated. The crude material was purified by flash chromatography (5-10% MeOH/CH₂Cl₂) to yield 1.9 g product (20% yield). ¹H NMR (DMSO-d₆) 7.80 (t, 1H), 4.02 (q, 2H, J=7.1Hz), 3.25-3.60 (m, 8H), 3.05 (q, 2H, J=6.0Hz), 2.58 (t, 2H, J=6.8Hz), 2.26 (t, 2H, J=6.3Hz), 1.57 (m, 4H), 1.47 (m, 4H), 1.15 (t, 3H, J=7.1Hz).

1.0 g of the amino ester (2.64 mmol) was treated with 1.1 g of biotin-NHS ester (3.22 mmol, 1.2 eq.) in 10 mL DMF with 1 mL of DIEA. The solution was allowed to stir overnight, and was then partitioned between brine (200mL) and ethyl acetate (200 mL). The ethyl acetate was dried, and the product purified by flash column chromatography (5-15% MeOH/CH₂Cl₂) to yield 870 mg (54%).

800 mg of the purified ester was suspended in 5 mL MeOH and treated with 5 mL of 1M KOH. The solution was stirred at 40 °C for 1 hour. The reaction mixture was acidified to pH 2-3 with concn. HCl then extracted four times with 25 mL Ethyl Acetate. The ethyl

acetate extracts were combined, dried (Na_2SO_4), and evaporated to give pure white product **4**, 640 mg (84% yield).

His₆ conjugate (**10**): The His₆ conjugate was prepared by linear solid phase peptide synthesis on PAM glycine resin. Following the six Histidine coupling performed using standard Boc chemistry, 3 Ado residues were added using Fmoc chemistry, followed by standard polyamide solid phase synthesis. The compound was cleaved from the resin with HF at the caltech peptide synthesis facility, and purified by HPLC. The compound gave a single peak after HPLC purification on an analytical HPLC run (C18, linear gradient from 0-60% CH_3CN in 0.1% TFA).

3.6 References

- (1) Honigberg, S. M.; Rao, B. J.; Radding, C. M. *Proc. Natl. Acad. Sci. USA* **1986**, 83, 9586-9590.
- (2) Rigas, B.; Welcher, A. A.; Ward, D. C.; Weissman, S. M. *Proc. Natl. Acad. Sci. USA* **1986**, 83, 9591-9595.
- (3) Ji, H. M.; Smith, L. M. *Anal. Chem.* **1993**, 65, 1323-1328.
- (4) Ito, T.; Smith, C. L.; Cantor, C. R. *Proc. Natl. Acad. Sci. USA* **1992**, 89, 495-498.
- (5) Beal, P. A.; Dervan, P. B. *Science* **1991**, 251, 1360-1363.
- (6) Moser, H. E.; Dervan, P. B. *Science* **1987**, 238, 645-650.
- (7) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, 382, 559-561.
- (8) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, 391, 468-471.
- (9) Ito, T.; Smith, C. L.; Cantor, C. R. *Nucl. Acids Res.* **1992**, 20, 3524-3524.
- (10) Nishikawa, N.; Oishi, M.; Kiyama, R. *J. Biol. Chem.* **1995**, 270, 9258-9264.
- (11) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; 2nd Ed. ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 1989.
- (12) Orum, H.; Nielsen, P. E.; Jorgensen, M.; Larsson, C.; Stanley, C.; Koch, T. *Biotechn.* **1995**, 19, 472-480.
- (13) Dervan, P. B. *Science* **1986**, 232, 464-471.
- (14) Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. *Science* **1989**, 243, 85-88.
- (15) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochem.* **1994**, 33, 3055-3062.
- (16) Iverson, B. L.; Dervan, P. B. *Meth. Enz.* **1993**, 218, 222-227.
- (17) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6141-6146.

Chapter 4:

γ -Aminobutyric Acid Analogs for Polyamide Hairpin Formation

4.1 Background

The discovery of the 2:1 polyamide motif provided critical information for the rational design of higher affinity DNA binding ligands. Improved affinity was essential to the efforts to use polyamides for the control of biological processes such as transcription. Common protein transcription factors have binding affinities for DNA on the order of nanomolar concentrations. Unlinked polyamide molecules did not begin to approach this goal. Mrksich and Dervan successfully investigated linking polyamide molecules either through an N to C terminal linkage, a “hairpin,”¹ or through an aliphatic linkage from the pyrrole ring nitrogen, an “H-pin.”^{2,3} The affinity hairpin polyamide ImPyPy- γ -PyPyPy-Dp increased by a factor of approximately 400 over the comparable unlinked homodimer resulting in an apparent binding constant of $8 \times 10^7 \text{ M}^{-1}$.^{1,4} This linkage provides substantial reduction in the entropic cost of bringing two polyamides to the binding site, while also controlling the binding mode of the polyamide. While this increase was quite favorable, it seemed as if a larger increase might be possible through further rigidification of the system. Not accounting for the cooperativity inherent to the system, it seemed possible to achieve closer to an additive energetic gain, or a compound with an affinity in the range of 10^{10} M^{-1} . Toward this end, Junyeong Cho synthesized the same polyamide cyclized through two γ -aminobutyric acid residues.⁵ This compound showed a further

increase in affinity of a factor of almost 40, but lost some of the specificity of the parent compound.

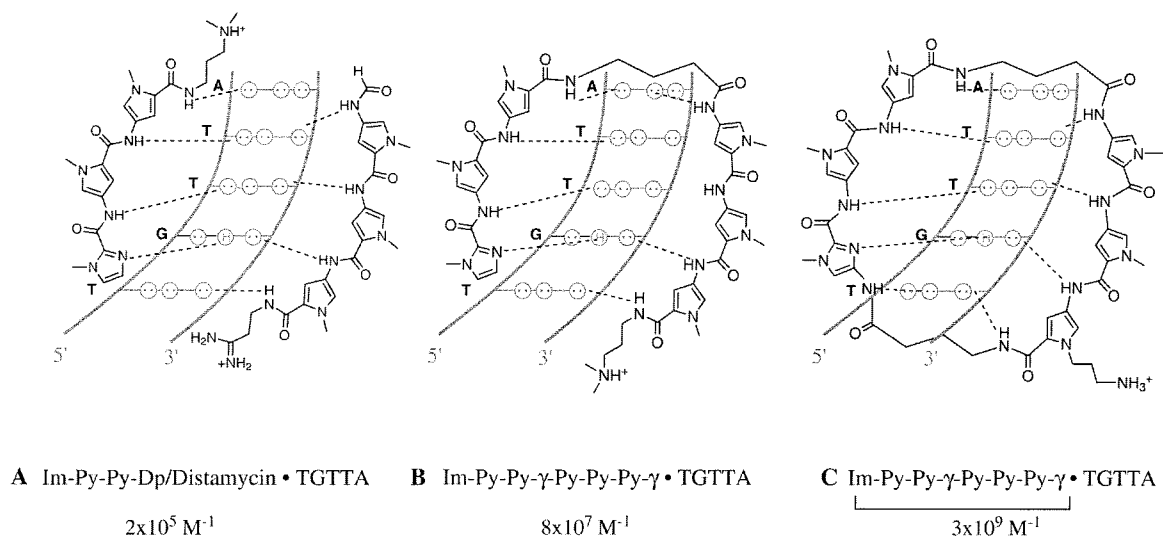


Figure 1. Improved binding of constrained hairpin polyamides.

Due to the relative synthetic difficulty of preparing cyclic polyamides rather than linear polyamides, we sought to design rigid, conformationally constrained analogs of γ -aminobutyric acid suitable for incorporation into polyamides. Using computer modeling, a series of analogs (shown in figure 2) were designed that approximated the calculated conformation of the γ -aminobutyric acid turn residue in a polyamide. The cis-cyclopropyl

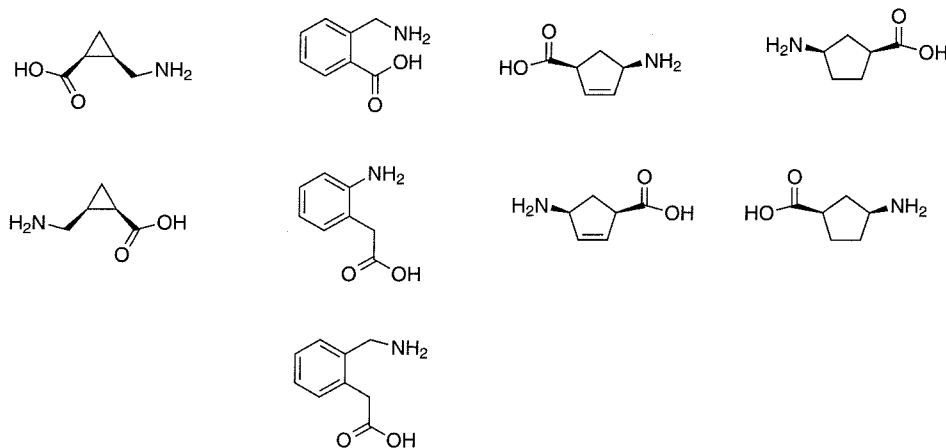


Figure 2. Constrained analogs of γ -aminobutyric acid

amino acids were studied by visiting Prof. N. Carballeira in collaboration with Eldon Baird and were not successfully incorporated into a polyamide. The rest of the turns were successfully synthesized as Boc protected monomers and, with the exception of the *o*-aminomethyl benzoic acid turn residue, they were incorporated into the same hairpin polyamide studied by Mrksich, Parks, Baird and Cho: ImPyPy-TURN-PyPyPy- β -Dp. The affinity and specificity of compounds were characterized by DNase I footprinting.

Subsequent to these initial efforts to synthesize rigid constrained γ -aminobutyric acid analogs, David Herman and Eldon Baird showed that (R)- α,γ diaminobutyric acid increased the affinity and specificity of the same hairpin.⁶ This commercially available amino acid derivative has proven to be extremely useful for the preparation of polyamide conjugates, for increasing the solubility of longer polyamides, and for providing a positive charge for compounds reductively cleaved from the resin.⁷ In spite of this success, the position of the α amine in this compound did not appear optimal for all conjugation experiments. Preliminary experiments showed that placement of a bromoacetyl group on the α turn position, for instance, did not result in DNA alkylation.⁸ Modeling suggested that a β,γ -diaminobutyric acid residue could bind DNA and place an amine along the floor of the minor groove of DNA, potentially a more suitable point of attachment for intercalators or for groups that interact or react with the floor of the minor groove. The preliminary modeling clearly identified the S enantiomer as the desired enantiomer, as the R is clearly sterically disfavored. Modeling also suggested that the S-amino isomer may also clash with the floor of the groove. Energetically, this energetic penalty might be compensated through formation of a favorable hydrogen bond to the A or T at the turn position in the binding site. To study this novel chirally substituted turn residue, it was synthesized and incorporated into the same parent hairpin polyamide structure, ImPyPy-TURN-PyPyPy- β -Dp, and is being studied by quantitative DNase I footprinting by David Herman.

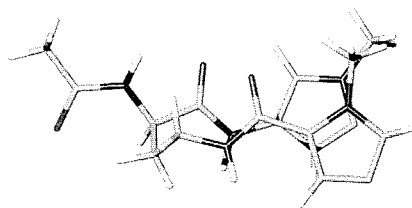
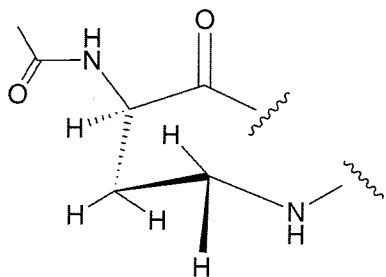
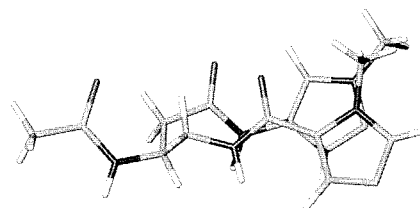
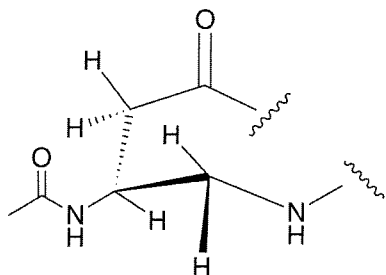
(R) α,γ diaminobutyric acid(S) β,γ diaminobutyric acid

Figure 3. Comparison of α,γ and β,γ diaminobutyric acid turn residues. Three-dimensional models were developed from NMR coordinates of a hairpin polyamide.⁹

4.2 Synthesis of Protected Turn Amino Acids

Synthesis of the 1,3 cis cyclohexyl derivatives was facilitated by the commercial availability of the optically active Vince's Lactam, 2-Azabicyclo[2.2.1]Hept-5-En-3-One, **1** and **7**. The lactam could be hydrolyzed and protected to yield the Boc derivatives, **3** and **9**. Alternatively, the lactam could first be hydrogenated with palladium on carbon as a catalyst, followed by hydrolysis and protection to yield compounds **6** and **12**.¹⁰

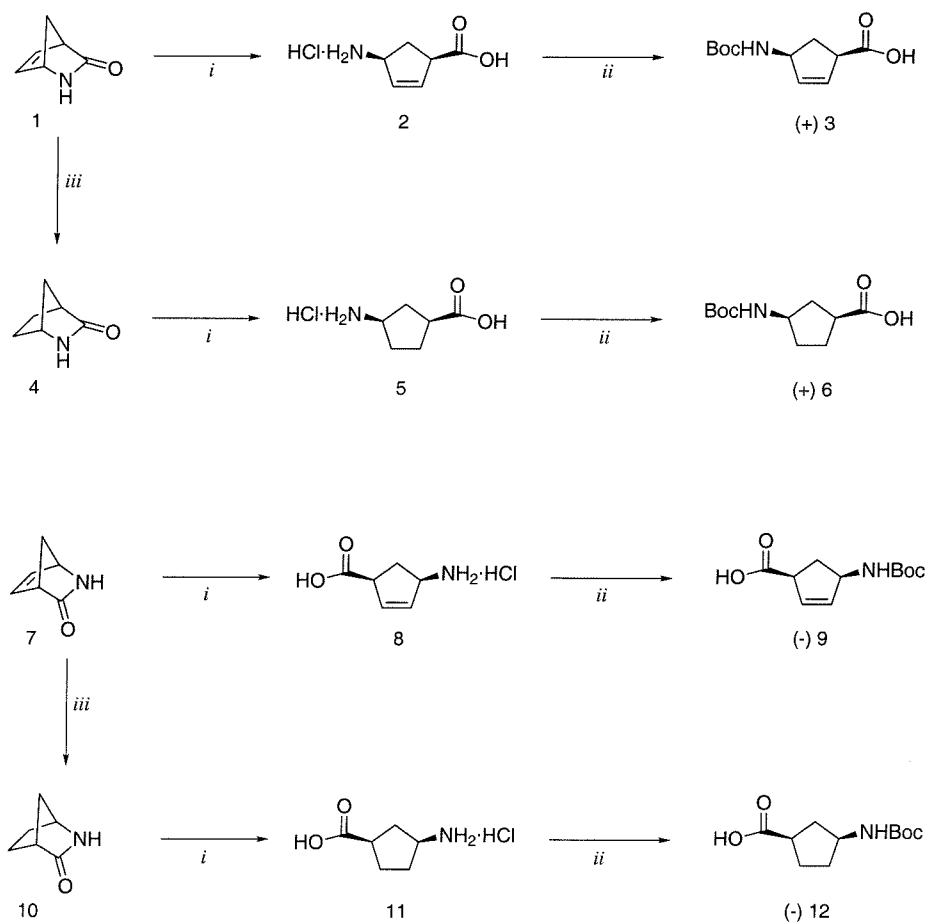


Figure 4. Synthesis of cyclopentyl amino acids. i) concn. HCl, 85-95%; ii) Boc₂O, NaOH 50-90%; iii) H₂, Pd/C 80-90%.

The *ortho* Boc-amino phenylacetic acid derivative 16 was readily synthesized from the commercially available nitro acid by esterification, hydrogenation and protection, followed by saponification of the ester, giving a 69% yield over three steps.

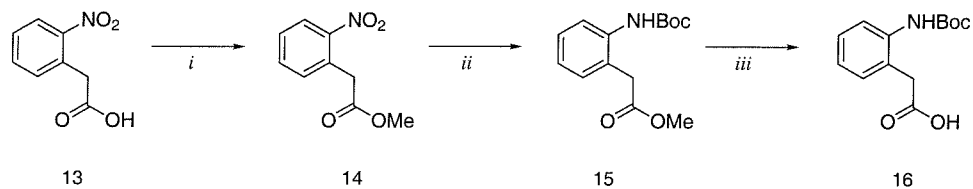


Figure 5. Synthesis of 2-Boc-amino-phenylacetic acid. i) MeOH, HCl, 98%; ii) Boc₂O, DMF, DIEA, 73%; iii) KOH, MeOH, H₂O, 96%.

Synthesis of the aminomethyl phenylacetic acid derivative 19 made use of the published Schmidt ring expansion of the available ketone 17 to provide lactam 18 in modest yield. This lactam was opened with acid hydrolysis and the crude amino acid protected to give protected amino acid 19, in moderate yield.¹¹ This sequence was not optimized as the initial synthesis provided sufficient material to continue.

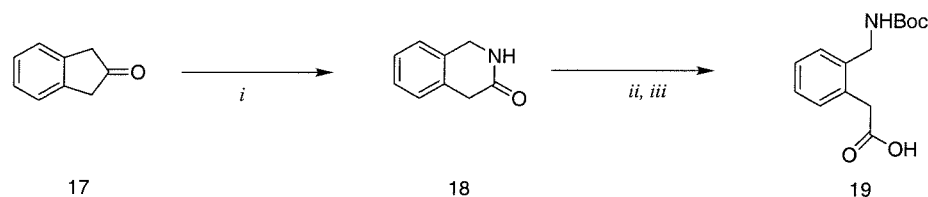


Figure 6. Synthesis of *o*-Boc-aminomethylphenylacetic acid. *i*) NaN_3 , H_2SO_4 , 88%; *ii*) concn. HCl ; *iii*) Boc_2O , diisopropylethylamine, DMF, 65%.

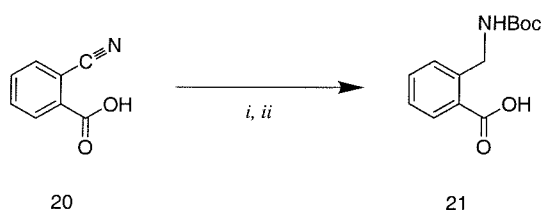


Figure 7. Synthesis of *o*-Boc-aminomethylbenzoic acid. *i*) H_2 , Pd/C; *ii*) Boc_2O , diisopropylethylamine, 12%.

Synthesis of polyamides containing aminomethyl benzoic acid derivative 21 proved to be the most difficult. A practical synthesis of 21 resulted from the reduction of the available *o*-cyanobenzoic acid by hydrogenation over platinum oxide followed by Boc protection of the resulting amine, to furnish the desired product in poor yield. Unfortunately, polyamide synthesis with this monomer failed repeatedly. Several conditions for activating the acid were attempted including activation with dicyclohexylcarbodiimide (DCC) and hydroxybenzotriazole (HOBt), with HBTU, or through preformation of a symmetric anhydride with DCC. It seemed likely that this compound might be too sterically hindered to couple efficiently on the solid support. To

avoid this coupling, the dimeric compound 25, was prepared to avoid this. Coupling the *o*-cyanobenzoic acid to a pyrrole residue proceeded smoothly. Reduction of the cyano group under improved conditions using cobaltous chloride hexahydrate with sodium borohydride resulted in good conversion to the amine.¹²⁻¹⁴ Boc protection yielded the fully protected amino acid in good yield. Saponification of the dimer provided the desired product in almost quantitative yield. Unfortunately, although this compound could be coupled to a growing polyamide on the solid support, further chain extension was not possible. This is similar to the result observed for the cyclopropyl derivatives studied by Carballeira and Baird. It seems possible that under the coupling conditions, the aminomethyl compound cyclizes to give a hemiacetal like intermediate incapable of further coupling, although this was not observed during the Boc protection of compound 23.

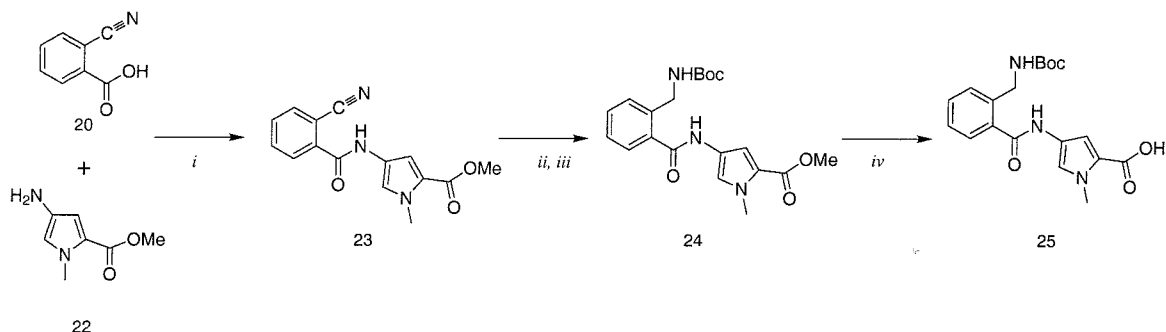


Figure 8. Synthesis of aminomethyl benzoic acid-pyrrole dimer. *i*) DCC/HOBt, DMF, 72%; *ii*) $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, NaBH_4 , MeOH; *iii*) Boc_2O , DIEA, DMF, 62%; *iv*) KOH, MeOH/ H_2O , 97%.

Synthesis of a protected β,γ -diaminobutyric acid for hairpin polyamide formation proceeded smoothly using the published route of Tomasini to the protected amino acid 30 as a guide.¹⁵ Crystallization of the S,S diastereomer 29 from cyclohexane:ethyl acetate afforded pure white crystalline material. This cyclic urea was opened in refluxing hydrochloric acid to afford the tosyl protected amino acid 30. Boc protection of the crude product gave the γ -Boc, β -tosyl protected acid in good yield. This amino acid was incorporated into a polyamide by solid phase synthesis. Incorporation of the turn was

confirmed by MALDI-TOF mass spectrometry ($M+H^+$ calcd. 1146.50 found 1146.68). The tosyl protecting group proved difficult to remove. In particular, the most likely deprotection method using HBr/HOAc/Phenol led to efficient polyamide degradation at room temperature or at elevated temperatures.^{16,17} In light of this failure, we decided to change protecting groups on the amino acid prior to solid phase synthesis. γ -Boc, β -tosyl protected acid 31 was smoothly reduced under Birch conditions and reprotected with Fmoc-OSu after a brief aqueous workup to give the desired (S)- γ -Boc- β -Fmoc-diaminobutyric acid. This compound was incorporated into growing polyamides with DCC/HOBt activation in good yield.

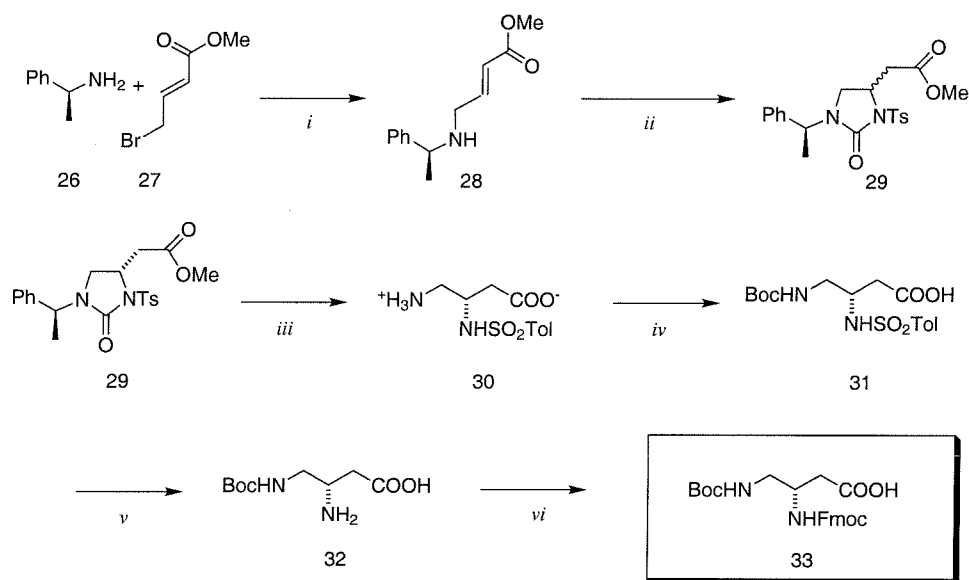


Figure 9. Synthesis of β,γ -diaminobutyric acid derivative, 33. *i*) DMF, 65°C, 85%; *ii*) tosylisocyanate, >95% conversion, 24% recovery of 29; *iii*) concn. HCl; *iv*) Boc₂O, NaHCO₃, dioxane/H₂O, 57%; *v*) Na/NH₃, THF; *vi*) Fmoc-OSu, Na₂CO₃, dioxane, H₂O, 47%.

4.3 Results and Discussion

With the exception of the aforementioned aminomethyl benzoic acid derivative 21, all of these compounds were readily introduced into hairpin polyamides by solid phase polyamide synthesis.¹⁸ The polyamides synthesized are shown in figure 10. The

polyamides were purified by preparative HPLC to give >95% purity by analytical HPLC.

The polyamides were each characterized by mass spectrometry and by NMR spectroscopy.

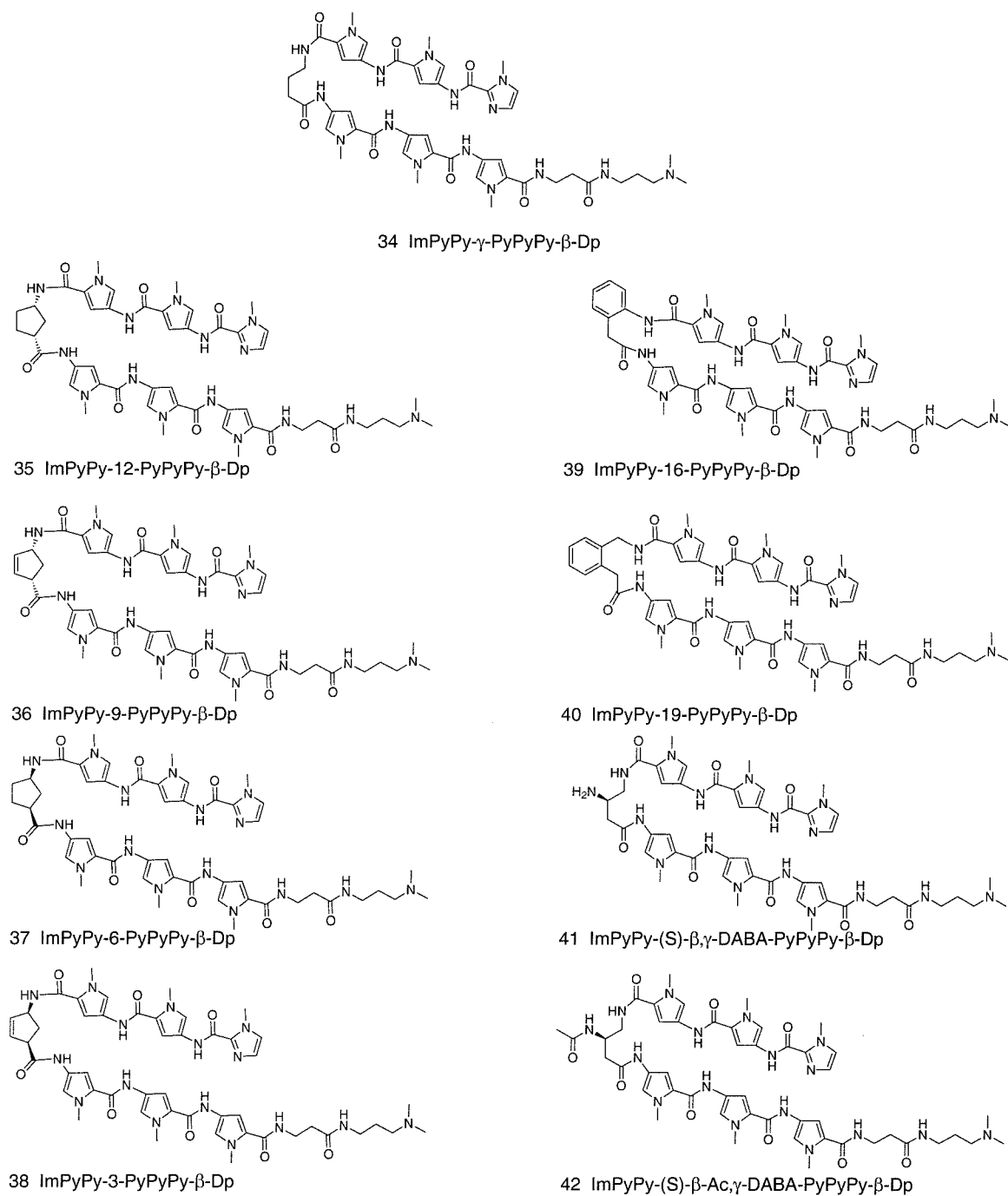


Figure 10. Polyamides used in this study.

Compounds were footprinted by quantitative DNase I footprint titrations on plasmid pMM5 that contains both 5'-TGTTA-3' match and 5'-TGACA-3' single base mismatch binding sites. Representative gels from the footprint titration of the *cis*-cyclopentyl amino acid containing polyamides **36** and **37** are shown in figure 11. Binding constants for each compound were determined by published methods,¹⁹ and are summarized in figure 13. Although most of the constrained compounds resulted in decreased binding affinity for the match DNA site, compounds **36** and **37** display stereospecific binding, with one enantiomer binding greater than 10 fold more tightly than the other.

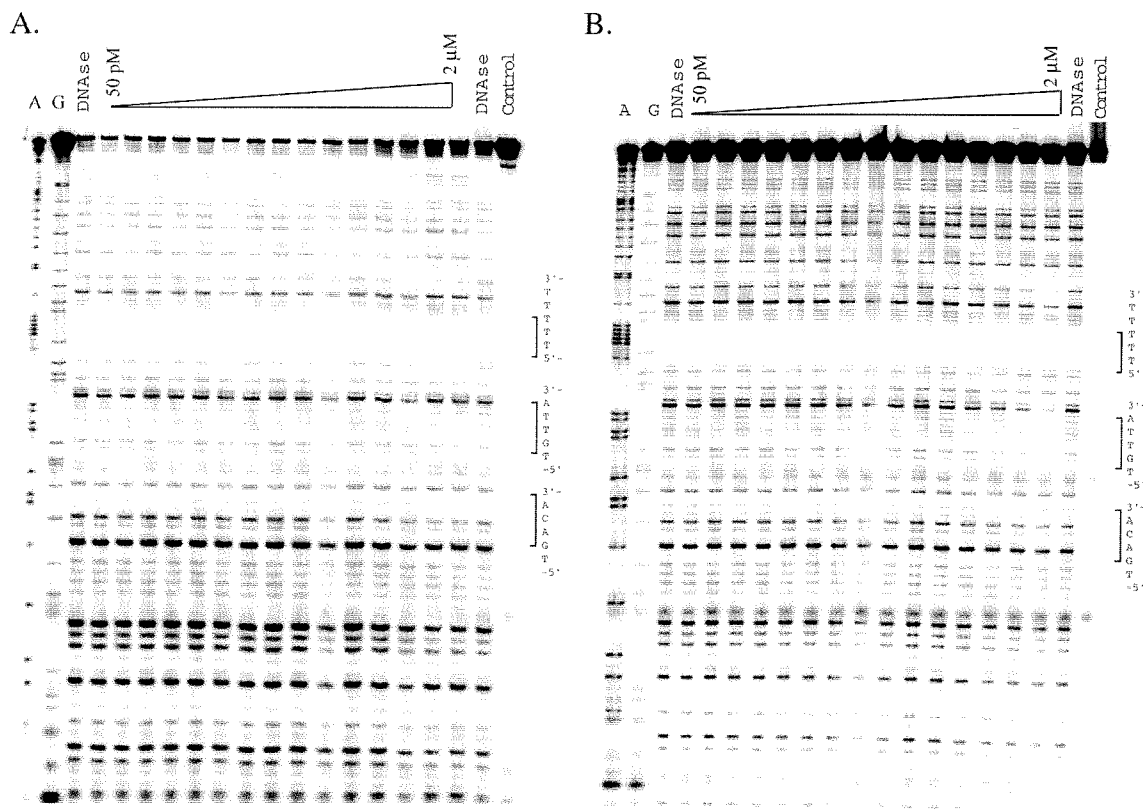


Figure 11. DNase I footprinting of polyamides 37 (A) and 35 (B). Footprint titrations were performed in 40 μ L volumes with final buffer conditions as follows: 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, and pH 7.0. Polyamide was equilibrated with EcoRI/BsrBI fragment of plasmid pMM5¹ for 12-18 hours prior to reaction with DNase I. Lanes marked DNase contain no polyamide. A and G lanes are chemical sequencing reactions.²⁰

Contrary to our goal, all of the constrained turns resulted in reduced affinity from the parent γ -aminobutyric acid linked compound. The least constrained of the compounds containing the 2-aminomethyl phenylacetic acid derivative shows an affinity comparable to the unconstrained γ -aminobutyric acid derivative and a comparable specificity, but with significantly less atom economy. This compound could still be of some interest for studying the control of polyamide binding modes. The remaining compounds all showed significant decreases in affinity and reduced specificity indicating that the conformational rigidity interfered with the preferred binding mode.

Computer modeling of these turns was a very useful tool for predicting the results of this study. At the time of this study, no structures were available of a hairpin polyamide. Using simple molecular mechanics and dynamics calculations, a turn was optimized that almost completely matched that determined later by NMR spectroscopy.⁹ A simple computational strategy of comparing the strain energy of a turn's predicted low energy conformation with that in the hairpin binding mode successfully predicted the rough order of binding affinities within the cyclopentyl class of compounds. Unfortunately, the approach was not able to account for steric interaction with the DNA, nor was it able to provide reliable energetic comparisons between systems, a common problem in computational chemistry. The developed models, even for the best optimized compounds, did show significant steric contact with the walls of the minor groove.

Since the completion of this work, more structural information has become available of polyamide interactions with DNA. Wemmer and coworkers determined the structure of a hairpin polyamide by NMR.⁹ In addition, the structures of several unlinked polyamides complexed with DNA have been determined at high resolution by x-ray crystallography.^{21,22} The hairpin structure is superimposable on the calculated gaba linked structure used in our design efforts. The gaba turn forms a particularly tightly folded

structure, which reasonably minimizes gauche interactions. The compact structure of the gaba turn in the bound structure is evidenced by the short 3.29Å distance between the carbon of the turn carbonyl and the nitrogen of the turn's amide linkage. In comparison, a simple cyclopentyl compound separates these atoms by 5.10Å in the favored diequatorial conformation and by 3.83Å in the higher energy diaxial conformation. Thus, in order to bind effectively, the cyclopentyl must distort past the high energy diaxial conformer.

While the designed turns can be forced to adopt similar conformations to the observed gaba conformation, it is likely that the introduced strain energy that precludes high affinity binding. The tighter binding isomers of the cyclopentyl series bind with similar structure to the gaba turn, placing a single methylene towards the floor of the groove. The enantiomers that place two carbons towards the groove are too bulky to bind efficiently. Both compounds suffer from projecting their pendant polyamide arms away at angles too obtuse for efficient binding. It could not be predicted with enough accuracy before the experiment what the magnitude of this effect would be on the overall binding energy.

The unlinked crystal structures may also provide some insight into the preference for a flexible turn residue over highly constrained molecules.^{21,22} The register of overlap between polyamide molecules bound in a side by side fashion differs in the structures solved. While there is undoubtedly some motion along the floor of the groove in the bound structure in solution, the differences between the crystal structures suggest that the precise relative orientations of the side by side polyamides will differ depending on sequence context. With this in mind, a more flexible linker may be more able to adopt a range of conformations that allows for high affinity binding when used in multiple polyamide motifs. More specifically, using the metric described above, a gaba turn would have to cover ~5.3Å in the ImImPyPy-β-Dp structure, ~4.2Å in the ImHpPyPy-β-Dp structure, while only 3.6Å in the structure of ImPyPyPy-β-Dp. This remarkable conformational

variability may require a turn residue that has several degrees of freedom rather than a compound that is fully rigidified.

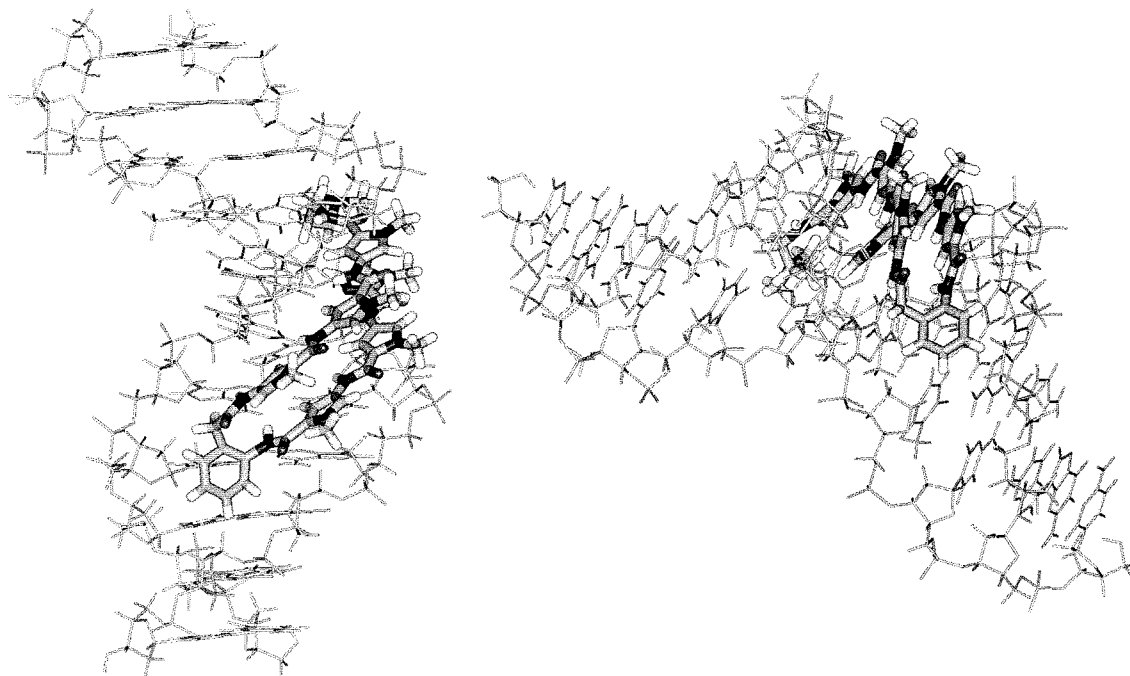


Figure 12. Computer model of constrained γ analog 39.

These studies support other work in the group showing that the gaba turn is almost optimal for the binding of hairpin polyamides, supporting the findings of Mrksich and Dervan that γ was superior to β as well as aminovaleric acid.¹ Subsequent to this work, Herman and Baird have studied a series of amino substituted γ turns, showing an increase in binding affinity for compounds bearing a chiral (R)-amino substituent at the α -position.⁶ Such modest modifications seem to be the limit of the motif's ability to withstand modification. It will be interesting to compare the (S)- β,γ -diaminobutyric acid turn moiety with the (R)- α,γ diaminobutyric acid.

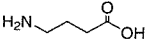
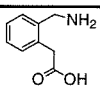
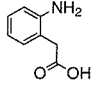
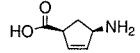
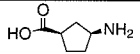
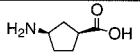
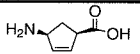
Polyamide	Turn	5'-TGTTA-3'	5'-TGACA-3'
34 ²³		2.9×10^8	4.8×10^6
40		1.0×10^8 (1.0)	4.5×10^7 (2.0)
39		9.7×10^6 (0.8)	1.2×10^6 (1.8)
36		1×10^6	$< 5 \times 10^5$
35		1×10^6	$< 5 \times 10^5$
37		$< 1 \times 10^6$	$< 1 \times 10^6$
38		$< 1 \times 10^6$	$< 1 \times 10^6$

Figure 13. Summary of quantitative footprinting results. All data were collected under standard conditions: 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, and pH 7.0. Polyamide was equilibrated with EcoRI/BsrBI fragment of plasmid pMM5¹ for 12-18 hours. Parentheses indicate standard deviations for three experiments.

4.4 Conclusions

Attempts to develop conformationally constrained γ -aminobutyric acid analogs that promote polyamide hairpin formation in the minor groove of DNA were largely unsuccessful. These analogs provide evidence that the γ linker is nearly optimal for hairpin formation. As solid phase techniques are further developed, it seems feasible that the fully constrained cyclic polyamides may become more important reagents. Additionally, since this project was completed, the range of affinities attained with polyamides has increased, mainly through increasing the length of the polyamide and its recognition site. As a result, further constraint may not be important for future recognition efforts. The relative success of polyamide 40 shows that there is some tolerance within the motif for modification. Formally, compound 40 is a constrained aminovaleric acid analog. As such it shows the designed improvement over the parent aminovaleric compound studied by Mrksich.¹ In

addition, this work represents the first study of chiral hairpin polyamides that bind DNA in a stereospecific manner. In light of the observed conformational variability in complexes between DNA and polyamides studied both by NMR spectroscopy and x-ray crystallography, it seems likely that a flexible turn unit may be better suited to allowing the polyamide to find its optimal arrangement with the DNA.

The β,γ -diaminobutyric acid turn has not yet been studied in a polyamide. While initial modeling suggests that the compound will suffer some energetic penalty for binding, it is hoped that this derivative will prove useful for further conjugation experiments. The synthesis of this compound is fairly simple, and the placement of the amine along the floor of the minor groove should make it a useful residue for the synthesis of conjugates of polyamides with alkylators, with minor groove binding compounds, or for the synthesis of tandem hairpins of the type under study by Herman and Baird. It is also expected that the recognition properties of this amino acid residue could be interesting. The polyamide places an amino group toward the floor of the groove such that it might be capable of forming a specific hydrogen bond with either A or T. In addition, the steric requirements should also increase the turn's specificity against binding to G-C or C-G base pairs. Given the recent success in recognition with hydroxypyrrole and hydroxylated β -alanine residues, it would also be interesting to characterize the γ -aminobutyric acid β -hydroxy (GABOB) derivative. GABOB can be readily synthesized by a variety of routes.²⁴⁻²⁷

4.5 Experimental

***N*-Boc cis-3-aminocyclopentenecarboxylic acids, 3 and 9.** 2 g (18.3 mmol) optically active 1 or 7 (98.5 and 99.5% ee respectively, Chiroscience Ltd., UK) were suspended in 8 mL of 4M HCl and stirred for 48 hours, then evaporated to give hydrochloride salts 2 and 8. The crude salts were dissolved in 18 mL of 2.5M NaOH + 35 mL *t*-butanol. 4.91 g (22 mmol, 1.2 eq.) of di-*t*-butyldicarbonate was added and the mixture stirred at room temperature for 1.5 hours. 45 mL of water was added, and the solution was washed with 2 x 40 mL of hexanes. The solution was then acidified with 6M HCl to approximately pH 2 and extracted 3 times with 40 mL of ethyl acetate. The combined ethyl acetate fractions were dried and evaporated. Recrystallization from ethyl acetate yielded pure products **3** (2.4 g, 58%) and **8** (2.3 g, 55%). ¹H NMR (DMSO-*d*₆) 1.38 (s, 9H), 1.71 (m, 1H), 2.40 (m, 1H), 3.40 (m, 2H), 4.47 (m, 1H), 5.71 (d, 1H), 5.81 (d, 1H), 6.92 (d, 2H, *J*=7.8 Hz).

***N*-Boc cis-3-aminocyclopentanecarboxylic acids, 6 and 12.** 2 g (18.3 mmol) of optically active lactams 1 and 7 were treated with 2 g of 10% Pd/C and hydrogenated under 1 atm of H₂ for 4 days, and filtered through celite to give pure lactams 4 and 8. Lactams 4 and 8 were treated exactly as describe above to give boc protected amino acids **6** (2.9 g, 69%) and **12** (2.1 g, 50%). ¹H NMR (DMSO-*d*₆) δ 1.36 (s, 9H), 1.61 (m, 1H), 1.73 (m, 2H), 2.05 (m, 2H), 2.63 (m, 1H), 3.71 (m, 1H), 6.84 (d, 2H, *J*=7.8 Hz).

methyl 2-nitro phenylacetate, 14. 10 g (55.2 mmol) of 2-nitro phenylacetic acid (13) were suspended in 250 mL anhydrous methanol. 5 mL of concn. H₂SO₄ was added and the mixture refluxed for 2 hours until all material is in solution. The volume was reduced by rotary evaporation and then diluted carefully with 100 mL of saturated sodium bicarbonate. The solution was extracted twice with 200 mL dichloromethane. The

combined dichloromethane extracts were washed with 50 mL of saturated NaHCO_3 then dried over Na_2SO_4 and evaporated to give pure ester 13, 10.4 g (53.2 mmol, 96%).

Boc-2-aminophenylacetic acid methyl ester, 15. 8 g (41.0 mmol) of 14 were dissolved in 50 mL DMF to which 0.4 g 10% Pd on carbon was then added. The mixture was hydrogenated for 12 hours under 350 psi of Hydrogen (23.8 atm). The resulting solution was filtered through celite and 16.0 g (73 mmol, 1.8 eq.) di-*t*-butyl dicarbonate was added along with 17.4 mL *N,N*-diisopropylethylamine. The reaction was stirred at 40°C for 12 hours, then partitioned between brine and ethyl ether (500 mL). The ether layer was dried and evaporated to give crude product. Chromatography using 4:1 hexanes: ethyl acetate to 3:1 hexanes: ethyl acetate afforded pure white product, 7.9 g (29 mmol, 73%). ^1H NMR (CDCl_3) δ 1.58 (s, 1H), 3.69 (s, 2H), 3.77 (s, 3H), 7.11 (t, 1H), 7.25 (d, 1H), 7.32 (t, 1H), 7.66 (br s, 1H), 7.82 (d, 2H).

Boc-2-amino phenylacetic acid, 16. 5 g (18.8 mmol) of ester 15 were dissolved in 200 mL of MeOH. To this solution was added 200 mL of 1M KOH (200 mmol, 10.6 eq.). The reaction was heated to 60°C for 1 hour. TLC using ethyl acetate showed conversion to a single spot. The reaction was extracted once with 200 mL of CH_2Cl_2 then acidified with 1M citric acid to pH 3-4. The solution was then extracted five times with 100 mL of dichloromethane. The organic phase was dried and concentrated to give 4.5g of pure acid. ^1H NMR ($\text{DMSO}-d_6$) δ 1.44 (s, 9H), 3.60 (s, 2H), 7.07 (t, 1H), 7.20 (t, 2H), 7.36 (d, 1H), 8.56 (s, 1H). ^{13}C NMR ($\text{DMSO}-d_6$) 28.2, 37.2, 39.5, 78.9, 124.5, 124.7, 127.1, 128.9, 130.9, 137.0, 153.5, 172.7.

***o*-Aminomethylphenylacetic acid δ -lactam, 18.** According to Naito et al.¹¹ 4.6 g (34.8 mmol) of 2-indanone (17) and 2.7 g (41.5 mmol, 1.2 eq.) of sodium azide were

dissolved in 7.5 mL of chloroform. To this solution, 7.5 mL of concn. H₂SO₄ was added slowly, maintaining the temperature between 30 and 40°C. The solution was then stirred at room temperature for 3 hours and poured onto 100 g of ice. The aqueous solution was extracted 3 times with 50 mL of ethyl acetate. The ethyl acetate was dried and evaporated to afford a thick brown oil. The pure product was obtained by recrystallization from ethyl acetate to yield 4.52 g (30.7 mmol, 88%). ¹H NMR (CDCl₃) δ 3.66 (s, 2H), 4.58 (s, 2H), 7.1-7.3 (m, 4H), 7.83 (br s, 1H).

***N*-Boc-2-methylaminophenylacetic acid, 19.** 2.17 g of lactam 18 (14.7 mmol) was refluxed in concn. HCl for 4 hours as previously described.¹¹ The reaction was filtered through a small amount of activated charcoal, then concentrated to give a thick oil. The oil was triturated with acetone overnight. The crude compound was then suspended in DMF (20 mL) to which diisopropylamine (5 mL) and Boc₂O were added. The solution was heated to 60°C for 4 hours. The reaction was partitioned between ethyl acetate and brine. The ethyl acetate layer was dried with Na₂SO₄ then evaporated to give crude product 19, that was then purified by flash chromatography using 5-10% MeOH in dichloromethane. ¹H NMR (CDCl₃) δ 1.40 (s, 9H), 3.52 (m, 2H), 4.24 (m, 2H), 7.1-7.4 (m, 4H).

Boc-2-aminomethylbenzoic acid, 21. 2.94 g of *o*-cyano benzoic acid (20.0 mmol) were dissolved in 100 mL of DMF. 300 mg of 10% Pd on carbon was added followed by hydrogenation under 250 psi H₂ (17 atm) for 14 hours. The reaction mixture was filtered through celite, then 11.0 g (50.4 mmol, 2.5 eq.) of Boc₂O was added. The reaction was stirred at room temperature for four hours, then partitioned between ethyl acetate and brine. The ethyl acetate was dried and evaporated to give crude 21, that was purified by flash

chromatography in 1:1 hexanes:ethyl acetate. ^1H NMR (CDCl_3) δ 1.44 (s, 9H), 5.31 (s, 2H), 7.25-7.6 (m, 4H).

Cyano benzoic acid- pyrrole dimer, 23. 5.0 g (34 mmol) of 2-cyanobenzoic acid (Transworld chemicals) was dissolved in 20 mL of peptide synthesis grade DMF. Hydroxybenzotriazole 4.6 g (34 mmol) and dicyclohexylcarbodiimide 7.0 g (34 mmol) were added, and the solution allowed to stand at room temperature for 3 hours. Separately, 7.85 g (42.5 mmol, 1.25 eq.) of methyl 4-nitropyrrole 2-carboxylate was hydrogenated in 20 mL DMF with 0.785 g of 10% Pd on carbon at 500 psi. After 4 hours, the reaction was filtered through celite. Diisopropylethylamine (11.0 g, 85 mmol, 2 eq.) was added. The activated cyano benzoic acid was then added by filtration to remove the dicyclohexyl urea. The reaction was then stirred overnight at room temperature. The reaction was partitioned between ethyl acetate and brine. The ethyl acetate layer was dried and evaporated. The residue was chromatographed on silica using from 3:1 to 1:1 hexanes:ethyl acetate to yield 6.9 g of pure product (24.5 mmol, 72%). ^1H NMR ($\text{DMSO}-d_6$) 3.35 (s, 3H), 3.76 (s, 3H), 7.30 (s, 1H), 7.60 (d, 1H, $J=1.3$ Hz), 7.72-7.85 (m, 3H), 8.21 (d, 2H, $J=7.4$ Hz), 10.40 (s, 1H).

Boc-2-methylamino benzoic acid pyrrole carboxamide methyl ester, 24.

1.13 g (4.0 mmol) of 23 and 0.948 g of $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ were dissolved in 15 mL of methanol. The solution was cooled on ice to 0°C . 0.757 g (20 mmol, 5 eq.) of sodium borohydride (NaBH_4) was added over 15 minutes. The reaction was followed by TLC using 1:1 hexanes: ethyl acetate. After 30 minutes, an additional 5 eq. of sodium borohydride was added, the reaction stirred an additional 15 minutes, then quenched with 15 mL of 3N HCl. The quenched solution was carefully neutralized with NaHCO_3 and extracted with ethyl acetate to give the crude amine. Amine: ^1H NMR (CDCl_3) δ 1.25 (t, 3H, $J=7.0$ Hz), 1.68 (s, 2H), 3.84 (s, 3H), 3.97 (s, 3H), 4.11 (q, 2H, $J=7.0\text{Hz}$), 5.66

(br s, 2H), 6.88 (s, 1H), 7.5-7.67 (m, 4H), 7.85 (d, 1H, $J=7.5\text{Hz}$). The crude amine was suspended in 10 mL of DMF to which is added 0.78 g (5.6 mmol, 2 eq.) diisopropylethylamine and 1.25 g (5.7 mmol, 2 eq.) of di-*t*-butyl dicarbonate (Boc_2O). The reaction was stirred at room temperature for 20 hours. After partitioning between ethyl acetate and brine, the evaporated product was purified by flash chromatography in 1:1 hexanes: ethyl acetate to give pure Boc protected dimer, 24. ^1H NMR (CDCl_3) δ 1.73 (s, 9H), 3.99 (s, 3H), 4.08 (s, 3H), 5.80 (d, 1H). ^{13}C NMR (CDCl_3) δ 27.32, 28.16, 36.64, 50.88, 57.89, 66.24, 80.85, 108.55, 119.98, 120.97, 123.06, 123.21, 129.46, 131.44, 132.18, 142.35.

Boc-2-aminomethylbenzoic acid pyrrole dimer, 25. 700 mg (1.8 mmol) of 24 were dissolved in 20 mL of MeOH. To this solution was added 20 mL of 1M KOH (5.5 eq.). The reaction was heated to 50°C for 3 hours, then extracted once with 25 mL ethyl acetate. The reaction was acidified to pH 3-4 with 1M citric acid, and extracted five time with 25 mL of ethyl acetate. The combined extracts were dried over Na_2SO_4 and evaporated to yield 656 mg (97%).

Allylic amine, 28. 6.1 g (50 mmol) of (S)-methylbenzylamine and 10.6 g (50mmol) of methyl bromocrotonate (85% pure, Aldrich) were dissolved in 100 mL of DMF and heated at 60°C for 12 hours. The reaction was partitioned between ethyl acetate and brine (250 mL each). The ethyl acetate was dried and evaporated to give crude 28 which was purified by silica gel chromatography using 2:1 hexanes: ethyl acetate to give 9.3 g (42 mmol, 85%). ^1H NMR (CDCl_3) δ 1.36 (d, 3H), 3.24 (d, 2H), 3.71 (s, 3H), 3.79 (q, 1H), 5.96 (dt, 1H), 6.94 (t, 1H), 6.97 (dt,), 7.2-7.4 (m, 5H). ^{13}C NMR (CDCl_3) 24.1, 47.9, 51.4, 57.6, 121.1, 126.5, 127.1, 128.5, 146.9, 166.7.

(5S,1'S) *N*-*p*-toluenesulfonylimidazolidin-2-one, 29. Allylic amine 28, 3.5 g (16.1 mmol), was treated with 3.17 g (16.1 mmol) of neat *p*-toluenesulfonylisocyanate at 0°C with rapid stirring. The solution solidified almost instantly to give a 1:1 mixture of 5R, 1'S and 5S, 1'S diastereomers in almost quantitative yield. The desired 5S material was crystallized repeatedly according to Tomassini *et al.*¹⁵ from 9:1 ethyl acetate: cyclohexane. The mother liquors contained the 5R isomer that was separable by careful chromatography on small scales. After repeated crystallization (2-3 times) the 5S diastereomer showed none of the 5R diastereomer by ¹H NMR. ¹H NMR (CDCl₃) δ 1.42 (d, 3H, *J*=7.1Hz), 2.44 (s, 3H), 2.47 (dd, 1H, *J*=10.2Hz, 17.0 Hz), 2.69 (dd, 1H, *J*=4.8Hz, 9.5Hz), 3.23 (dd, 1H, *J*=3.0 Hz, 16.8Hz), 3.62 (dd, 1H, *J*=7.4, 18.5 Hz), 4.52 (m, 1H), 5.15 (q, 1H, *J*=7.1Hz), 7.2-7.4 (m, 7H), 7.93 (d, 2H, *J*=8.2Hz).

(S)-*N*-β-toluenesulfonylamino-γ-Boc-aminobutyric acid, 31. 1.47 g (3.5 mmol) of the S,S diastereomer 29 were refluxed in 50 mL of concn. HCl for 40 hours until the suspension cleared. After cooling, the solution was extracted once with 50 mL of ethyl ether to remove the putative ethyl benzene byproduct. The acidic solution was then cooled on ice and neutralized with sodium hydroxide. To the neutral solution was added 5 g of NaHCO₃ followed by a solution of 1.16 g (5.3 mmol, 1.5 eq.) di-*t*-butyldicarbonate in 50 mL of dioxane. The solution was stirred overnight then carefully acidified to ~ pH 3 with 1M HCl. The solution was extracted with ethyl acetate that was then dried and evaporated to give the crude product. Flash chromatography on silica using 5-10% methanol in dichloromethane afforded the desired product, 747 mg (2.0 mmol, 57%). ¹H NMR (DMSO-*d*₆) δ 1.35 (s, 9H), 2.36 (s, 3H), 2.89 (m, 2H), 3.42 (m, 2H), 6.15 (d, 1H), 6.21 (t, 1H) 7.34 (d, 2H), 7.65 (d, 2H). ¹³C NMR (DMSO-*d*₆) δ 21.0, 28.2, 43.5, 50.8, 77.7, 77.8, 126.5, 129.5, 138.7, 142.4, 155.7.

(S)-N- β -Fmoc-amino- γ -Boc-aminobutyric acid, 33. 750 mg (2.0 mmol) of tosyl protected amine 31 were dissolved in tetrahydrofuran and added to a deep blue solution of sodium metal in liquid ammonia at -78°C . Sodium metal was added until the blue color persisted for at least 10 minutes. The solution was allowed to come to room temperature slowly and the ammonia allowed to evaporate. The resulting THF solution was diluted carefully with water and neutralized with hydrochloric acid. The THF was removed by evaporation leaving approximately 25 mL of solution. 2.5 g of sodium carbonate was added followed by a solution of succinimidyl 9-fluorenylmethyl carbonate (Fmoc-Osu) 1.35 g (4 mmol, 2 eq.) in 10 mL of dioxane. The reaction was stirred overnight. The pH was adjusted carefully to pH 3-4 and repeatedly extracted with ethyl acetate. The ethyl acetate was dried and evaporated to give crude material that was purified by silica gel chromatography with 5% methanol in dichloromethane. After drying under high vacuum, the product was obtained as 414 mg (0.94 mmol, 47%) of white powder. ^1H NMR ($\text{DMSO}-d_6$) δ 1.35 (s, 9H), 2.36 (m, 2H), 3.02 (m, 2H), 3.83 (m, 1H), 4.23 (m, 2H), 6.84 (t, 1H), 7.20 (d, 1H, $J=8.3\text{Hz}$), 7.31 (t, 2H, $J=7.4\text{Hz}$), 7.40 (t, 2H, $J=7.3\text{Hz}$), 7.67 (d, 2H, $J=7.3\text{Hz}$), 7.87 (d, 2H, $J=7.4\text{Hz}$). ^{13}C NMR ($\text{DMSO}-d_6$) δ 29.2, 47.4, 66.4, 78.8, 110.7, 121.1, 122.3, 126.2, 128.1, 128.6, 129.9, 141.7, 144.9, 156.5, 156.8, 173.5.

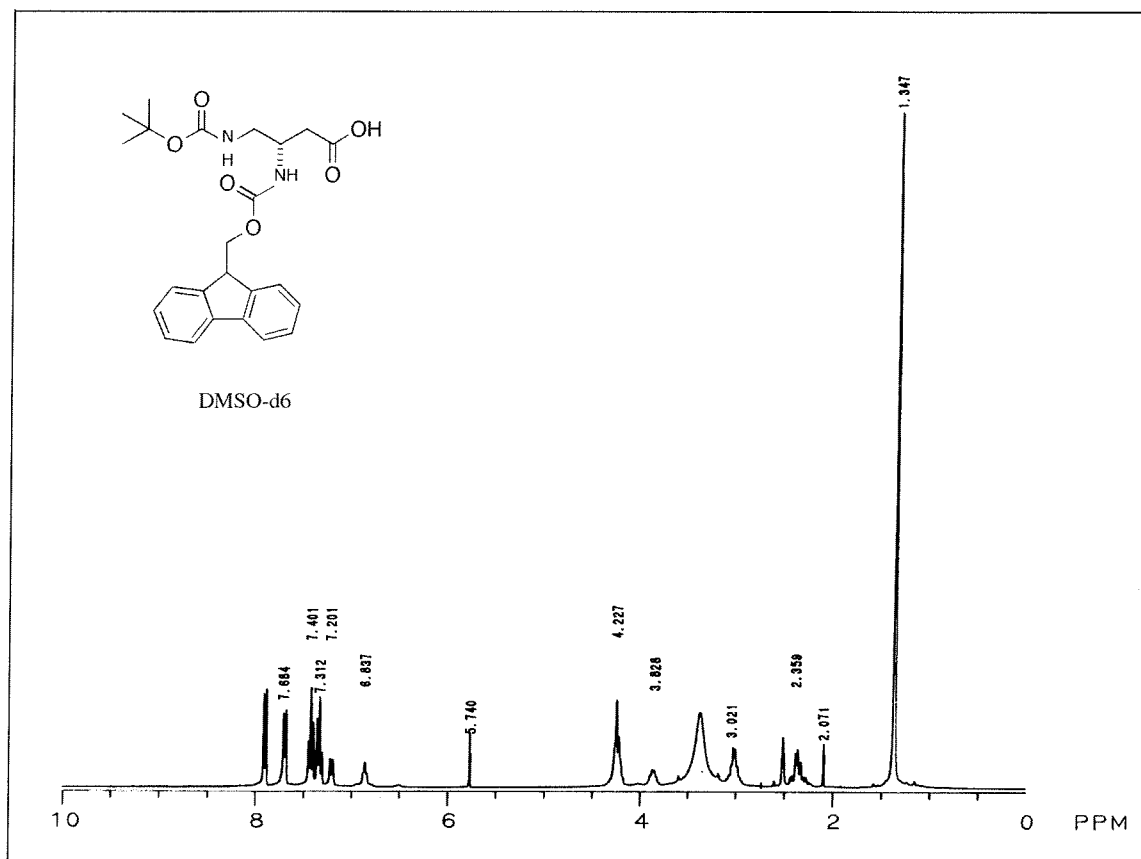


Figure 14. ¹H NMR of compound 33.

Polyamide Synthesis: Polyamides were synthesized according to the general procedures outlined by Baird and Dervan¹⁸ using β -alanine resin with a substitution level of 0.21 mmol/g. Each compound was synthesized on a 0.05 mmol scale. All boc deprotections were carried out with 100% trifluoroacetic acid. Turn amino acids were activated with DCC and HOBt in 1 mL of DMF and were used in four fold excess. The subsequent coupling to Boc-Py-OBt was also carried out with a four-fold excess. Polyamide synthesis was monitored by HPLC using a Rainin C18 analytical column and a gradient from 0 to 60% acetonitrile in 0.1% TFA/Water over 35 minutes. Compounds were cleaved from the resin with *N,N*-dimethylaminopropylamine and purified by preparative HPLC (Waters C18). The purified compounds were characterized by matrix

assisted laser desorption-time of flight mass spectrometry (MALDI-TOF) and by ^1H NMR with the exception of compounds 41 and 42 which were characterized only by MALDI-TOF and analytical HPLC. Mass spectrometry results for polyamides are as follows: **35** 1003.62 ($\text{M}+\text{H}^+$), calcd. 1003.52; **36** 1000.96 ($\text{M}+\text{H}^+$), calcd. 1001.49; **37** 1000.96 ($\text{M}+\text{H}^+$), calcd. 1001.49; **38** 1000.96 ($\text{M}+\text{H}^+$), calcd. 1001.49; **39** 1025.62 ($\text{M}+\text{H}^+$), calcd. 1025.49; **40** 1039.87 ($\text{M}+\text{H}^+$), calcd. 1039.49; **41** 992.50 ($\text{M}+\text{H}^+$), calcd. 992.49; **42** 1034.54 ($\text{M}+\text{H}^+$), calcd. 1034.50.

4.7 References

- (1) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.
- (2) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663-3664.
- (3) Chen, Y. H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6995-7005.
- (4) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576.
- (5) Cho, J. Y.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10389-10392.
- (6) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.
- (7) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Ang. Chem. Int. Ed.* **1998**, *37*, 1421-1423.
- (8) Mapp, A., personal communication .
- (9) deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909-7916.
- (10) Evans, C.; McCague, R.; Roberts, S. M.; Sutherland, A. G. *J. Chem. Soc. Perk. Trans. I* **1991**, 656-657.
- (11) Naito, T.; Okumura, J.; Kaisi, K.; Masuko, K.; Hoshi, H.; Kamachi, H.; Kawaguchi, H. *J. Antib.* **1977**, *30*, 698-704.
- (12) Satoh, T.; Suzuki, S.; Susuki, Y.; Miyaji, Y.; Imai, Z. *Tet. Lett.* **1969**, *52*, 4555-4558.
- (13) Heinzman, S. W.; Ganem, B. *J. Am. Chem. Soc.* **1982**, *104*, 6801-6802.
- (14) Osby, J. O.; Heinzman, S. W.; Ganem, B. *J. Am. Chem. Soc.* **1986**, *108*, 67-72.
- (15) Cardillo, G.; Orena, M.; Penna, M.; Sandri, S.; Tomasini, C. *Synlett* **1990**, 543-544.
- (16) Roemmele, R. C.; Rapoport, H. *J. Org. Chem.* **1988**, *53*, 2367-2371.

- (17) Haskell, B. E. *J. Org. Chem.* **1975**, *41*, 159-160.
- (18) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
- (19) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Meth. Enz.* **1986**, *130*, 132.
- (20) Iverson, B. L.; Dervan, P. B. *Meth. Enz.* **1993**, *218*, 222-227.
- (21) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147-6152.
- (22) Kolb, H. C.; Bennani, Y. L.; Sharpless, K. B. *Tet. Asymm.* **1993**, *4*, 133-141.
- (23) Lohray, B. B.; Reddy, A. S.; Bhushan, V. *Tet. Asymm.* **1996**, *7*, 2411-2416.
- (24) Bernabei, I.; Castagnani, R.; DeAngelis, F.; DeFusco, E.; Giannessi, F.; Misiti, D.; Muck, S.; Scafetta, N.; Tinti, M. O. *Chem. Eur. J.* **1996**, *2*, 826-831.
- (25) Misiti, D.; Zappia, G.; Dellemonache, G. *Synth. Comm.* **1995**, *25*, 2285-2294.

Chapter 5:

Recognition of G-Rich DNA Sequences

5.1 Background

The ability to specifically inhibit gene transcription with small molecules requires the ability to target all desired DNA sequences with affinities comparable to protein transcription factor binding affinities, generally in the nanomolar range. Pyrrole-imidazole polyamide molecules bind to DNA with affinities and selectivities comparable to protein transcription factors; however, some sequences remain challenging targets, either due to sequence context effects or mismatches in either the phasing or curvature of the polyamide molecules with respect to the DNA structure.^{1,2} Polyamides targeting three or more consecutive G residues bind with generally reduced affinity compared to those targeting mixed DNA sequences. Targeting such G rich sequences is essential for expanding the scope of the polyamide approach to *in vivo* gene regulation with small molecules.

Replacement of a pyrrole residue with β -alanine has been shown to restore high affinity binding for longer binding sites and for those containing multiple G residues. The polyamide ImPyImPy- γ -ImPyImPy- β -Dp, for example, shows relatively weak (4×10^7 M⁻¹) binding to its 5'-WGCGCW-3' target site. High affinity (4×10^9 M⁻¹) is obtained for this compound through the replacement of two pyrrole residues with β -alanine, for the compound, Im β ImPy- γ -Im β ImPy- β -Dp.³ This approach is not applicable for the recognition of homo G-rich sequences as one strand of the polyamide does not contain replaceable pyrrole residues. For the recognition of 5'-WGGGGW-3' sequences, shown

by compound **1**, ImImImIm- γ -PyPyPyPy- β -Dp, there is no possibility for β -alanine placement in the imidazole rich strand. This polyamide binds its target sequence selectively, but with an affinity of only $2.8 \times 10^7 \text{ M}^{-1}$ to a 5'-TGGGGA-3' site.⁴ Likewise, recognition of sequences containing three contiguous G residues also remains challenging. The polyamide ImImImPy- γ -PyPyPyPy- β -Dp recognizes a 5'-AGGGAA-3' site with good specificity but an equilibrium association constant of only $4 \times 10^8 \text{ M}^{-1}$.⁵ The goal of developing molecules that bind homo G sequences with subnanomolar affinity remains a challenge.

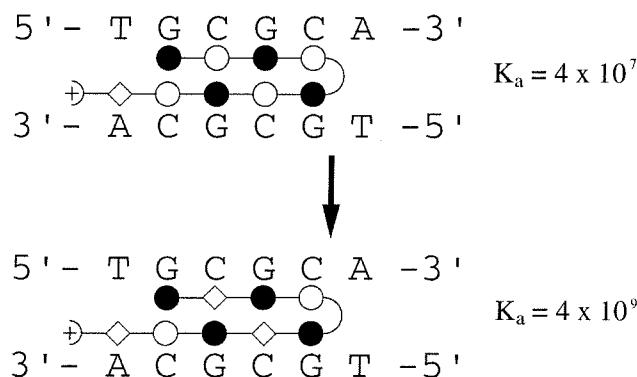


Figure 1. Restoration of high affinity binding through β -alanine linkers.³

A possible solution to this problem is the development of β alanine derivatives specific for guanine residues. This approach is currently being developed in the group by Paul Floreancig. While this approach is very rational, it has several limitations. First, β alanine residues in polyamides make close contacts with the floor of the minor groove making effective substitution sterically difficult. Second, imidazole residues seem particularly well suited to making hydrogen bond contacts with the 2-amino group of guanine.

Another possible approach to the recognition of G-rich sequences is the replacement of the rigid amide linkage between two imidazoles with a more flexible functional group.

Initial efforts have focused on the development of alkyl and amino linkages between two imidazole residues, though others are certainly plausible. While the alkyl approach formally removes a single hydrogen bond in the DNA complex, the large affinity increase observed using β -alanine replacement of pyrrole residues suggests that the increased flexibility may provide greater energetic benefits, balancing an entropic loss with an enthalpic gain through the formation of more optimized hydrogen bonds in the bound conformation. In addition, the loss of the single hydrogen bond occurs for both the match and mismatch complexes and should, therefore, not appreciably alter the specificity. The amine derivative was designed to retain the recognition, but with increased flexibility. The amine derivative is also more synthetically tractable, potentially increasing its practical utility as amines have been synthesized on the solid support by reductive amination.^{6,7}

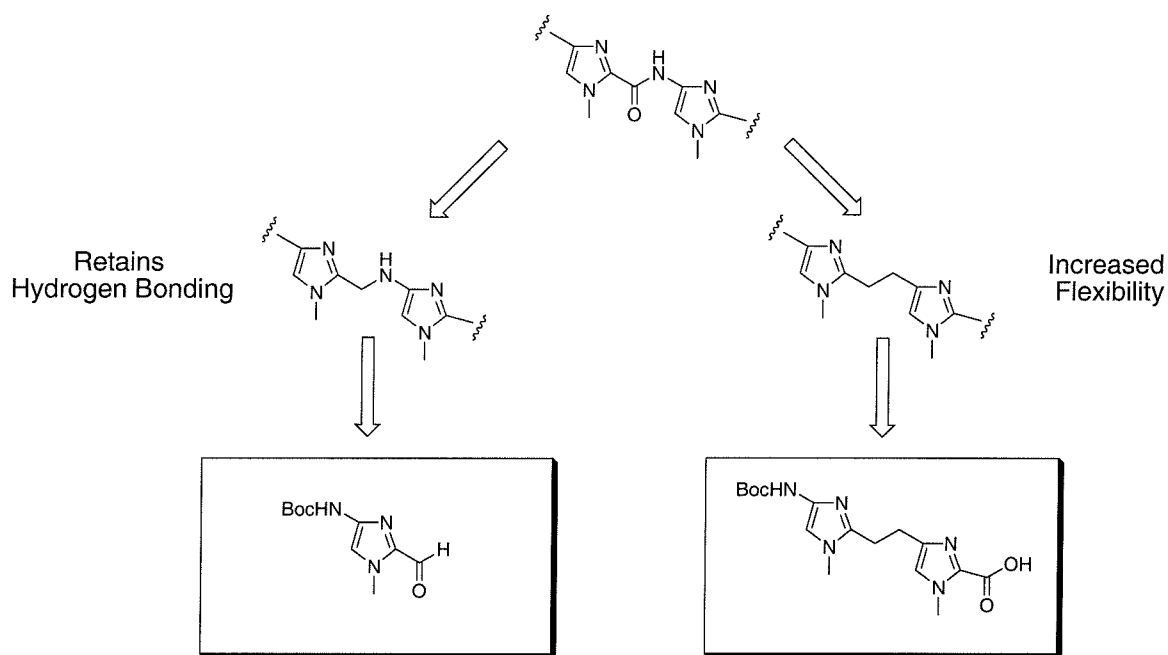


Figure 2. Replacement of amide linkage with conformationally flexible functional groups.

The relatively low affinity of the ImImImIm- γ -PyPyPyPy- β -Dp polyamide for its match site led us to choose this as a parent compound for this study. Four polyamides

were designed incorporating a flexible amide replacement between the central imidazoles, opposite either contiguous pyrroles, or pyrroles separated with a β -alanine residue. These compounds are shown in figure 3. This chapter focuses on the synthesis and characterization of alkyl derivatives **2** and **3**.

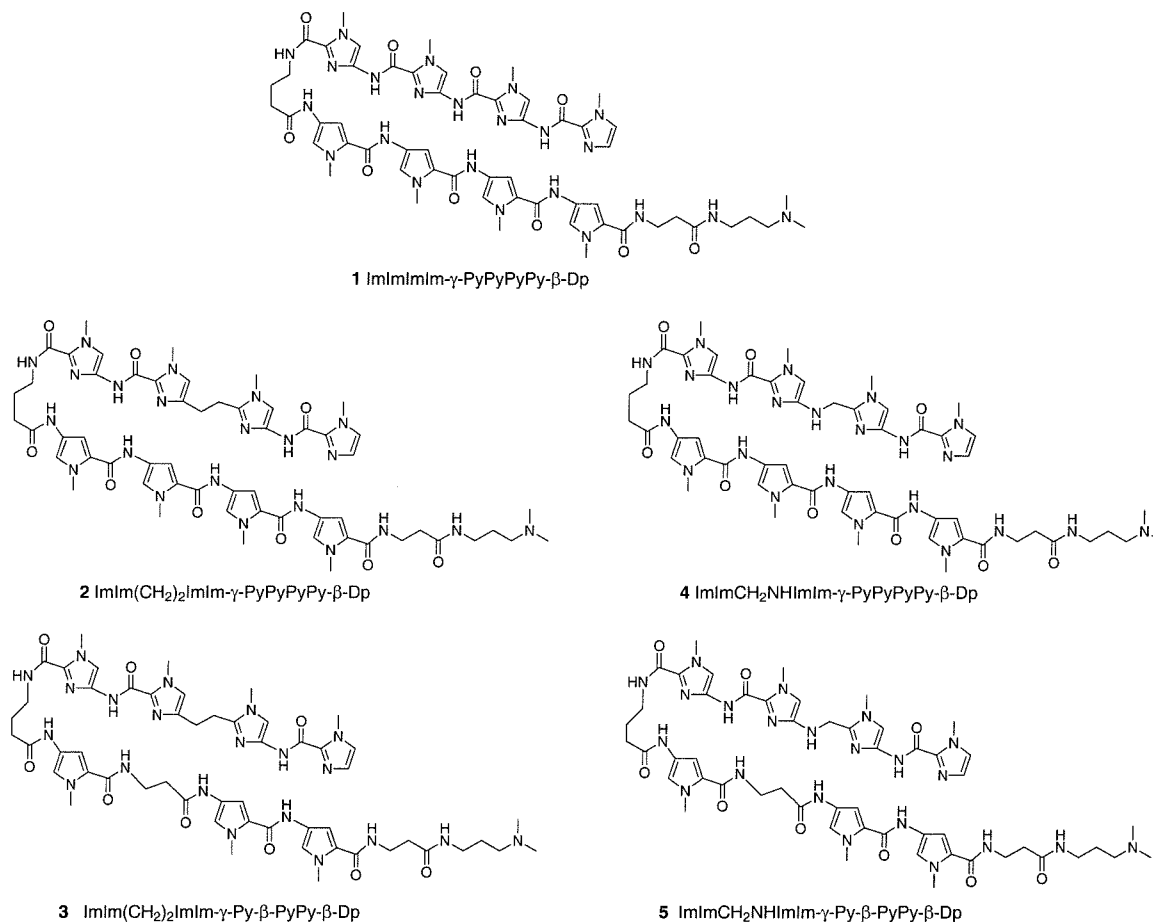


Figure 3. Polyamides designed to recognize 5'-WGGGGW-3' sequences.

5.2 Synthesis

The proposed compounds bearing alkyl linkages represented significant synthetic challenges, most importantly in defining the regiochemistry of the compounds. Early retrosynthetic analysis favored palladium mediated couplings as an attractive means of making such compounds, particularly in light of the synthesis of an alkene linked

distamycin⁸ analog using Heck couplings.⁹ While the Heck chemistry seemed attractive, literature reports suggested that the necessary vinyl imidazoles were highly toxic and unstable due to the propensity to polymerize.^{10,11} With this in mind, routes were explored that relied on the Sonagashira modification of the Castro-Stephens palladium coupling.^{12,13} The target alkyne dimer could be

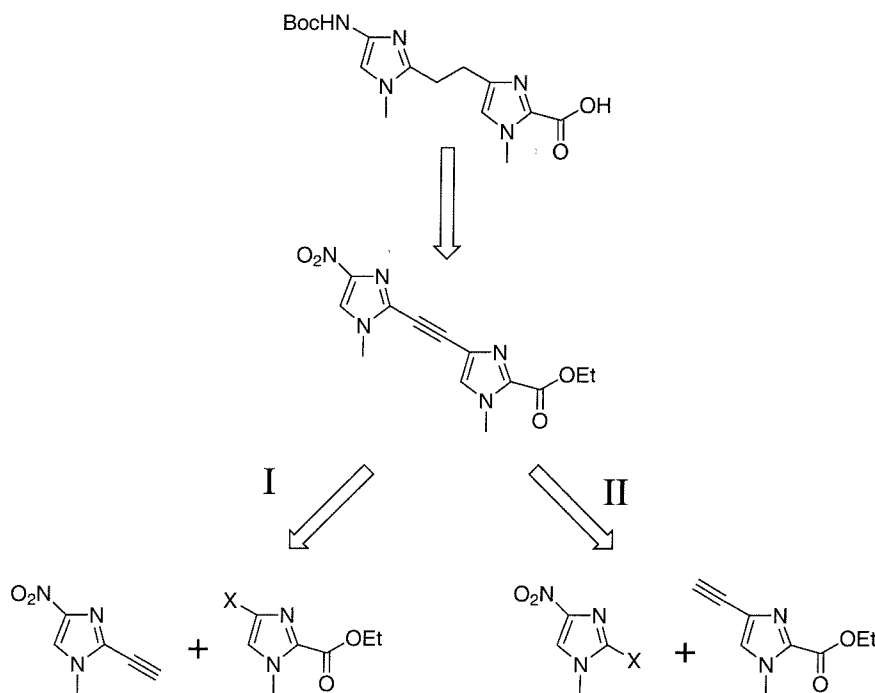


Figure 4. Retrosynthetic analysis of Boc protected alkyl diimidazole.

reached by two similar routes shown retrosynthetically in figure 4. Both routes were explored in small scale reactions and are viable, but route I was pursued more rigorously. The key intermediates shared in the pathways are 4-nitro 2-halo 1-methylimidazole, and 4-halo 1-methyl 2-ethylimidazole carboxylate. A second difficulty with this path was the synthesis of 1-methyl 4-nitroimidazole. Using the route of Benjes and Grimmett, acceptable regioselectivity could be achieved on small scales.¹⁴

The synthesis of the 1-methyl 2-halo 4-nitro derivatives provided some initial difficulty. The 1-methyl 2-bromo 4-nitro imidazole had been previously synthesized by an

inefficient route,¹⁵ but the 2-iodo derivative was unknown. Direct halogenation of 1-methyl-4-nitroimidazole with NBS, NIS, and I₂ all failed to cleanly give the desired product. Indirect halogenation following lithiation also failed to give the correct product, most likely due to lithium migration to the activated 5 position.

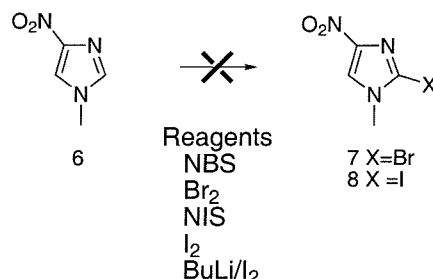


Figure 5. Failed halogenation of 1-methyl-4-nitroimidazole.

An important breakthrough in this synthesis came from the early observation of Lown and coworkers that 1-methyl-4-nitro-2-imidazole carboxylate is prone to decarboxylation.¹⁶ It was found that the well known compound could be directly converted to halogen derivative in the presence of NIS or NBS at elevated temperatures (70-100 °C). This reaction is a variation of the Hunsdiecker-Borodin reaction, normally proceeding through a silver salt of the carboxylate. Another group recently explored the metal free Hunsdiecker-Borodin reaction.¹⁷ Conversion of the readily synthesized carboxylate^{16,18,19} to the halo derivative solved one of the key regiochemistry problem in the synthesis.

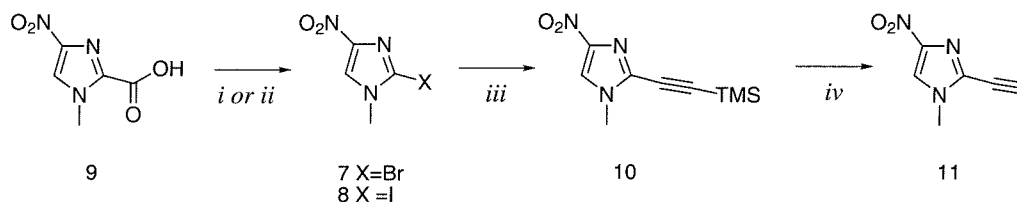


Figure 6. Modified Hunsdiecker-Borodin Sonagashira reaction sequence.

i) *N*-iodosuccinimide, toluene, DMF 80 °C; X=I, 60-75%; *ii*) *N*-bromosuccinimide, DMF 100 °C; X=Br, 91%; *iii*) trimethylsilylacetylene, Pd(PPh₃)₂Cl₂, DMF DIEA X=Br N.R.; X=I 18%, 2 steps; *iv*) NaOEt, THF 95%.

Sonagashira coupling of trimethylsilylacetylene (TMSA) to 4-nitro 2-bromo imidazole was unsuccessful, while coupling to the iodo derivative proceeded in reasonable yield, 18% over two steps. Synthesis of this compound was also achieved from the known compound 1-methyl 2-trimethylsilylacetylenoimidazole^{20,21} by nitration in fuming nitric acid at reflux in low yield, and production of both regioisomers.

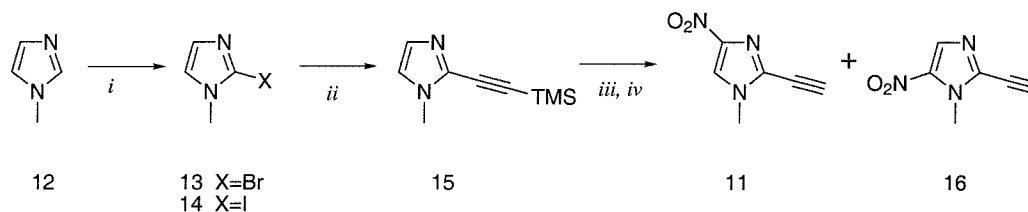


Figure 7. Synthesis of 1-methyl-4-nitro-2-acetylenoimidazole. *i)* BuLi, BR₂ or I₂; *ii)* trimethylsilylacetylene, Pd(PPh₃)₂Cl₂, DMF DIEA X=Br N.R.; X=I 40%; *iii)* HNO₃ reflux; *iv)* NaOEt, EtOH/THF.

Halogenation of 1-methyl-2-ethyl imidazole carboxylate with NBS provided the 4- and 5-bromo derivatives in moderate yield. Careful chromatography yielded pure 4-bromo derivative. As the 2-bromo imidazole **7**, this derivative was unreactive toward alkynylation. As it is known that iodo arenes are more reactive than bromo derivatives,²² the synthesis of the 4-iodo derivative was undertaken. No reaction occurred with 1-methyl-2-ethylimidazole carboxylate (**17**) and NIS. Iodination with I₂ resulted in the desired product as a mixture of 4-iodo and 5-iodo regioisomers, separable by repeated flash chromatography. Given the low yield and difficult chromatography, we sought an improved synthesis. Using more reactive iodination conditions, including addition of silver salts, resulted in higher yields but reduced selectivity, particularly in the production of the 4,5-diimidazole derivative **23** which interfered with chromatographic separation. Iodination of 1-methyl-2-trichloroketoimidazole at -30 °C in DMF afforded only a single isomer that could be converted with sodium ethoxide to the desired product, **19**.

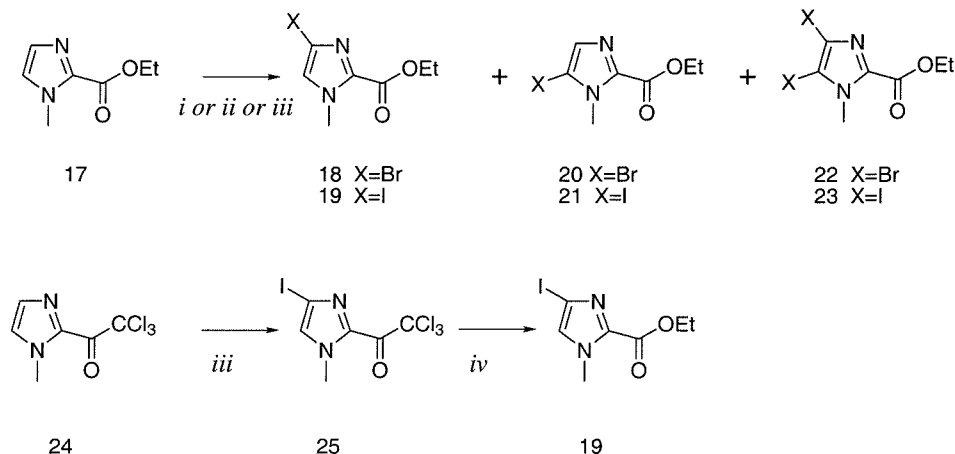


Figure 8. Synthesis of 4 halo imidazoles. *i)* *N*-bromosuccinimide, THF, reflux; *ii)* I_2 , DMF, 100 °C, 5 days; *iii)* I_2 /silver *p*-toluenesulfonate, DMF, -30 °C; *iv)* NaOEt/EtOH.

In order to establish the regiochemistry of the 4 and 5 iodo imidazoles, three approaches were undertaken. First, the compounds were assigned based on chemical shifts reported for analogous compounds lacking a 2-substituent.²³ In the reported compounds, H-4 always comes farther downfield in CDCl_3 . This was observed for compound **21**, 7.27 ppm versus 7.13 for the 4-isomer, compound **19**. Secondly, an NOE difference spectra, shown in figure 9, of a mixture of the two isomers with irradiation at the *N*-methyl group confirmed the assignment. A strong NOE was observed to H-5 with no NOE to H-4. Lastly, the improved synthetic route from the 2-trichloroketoimidazole **24** supports the regiochemical assignments as this compound undergoes electrophilic aromatic substitution reactions preferentially at the 4 position.¹⁹ All three analyses are consistent with the correct regiochemical assignment of compounds **19** and **21**.

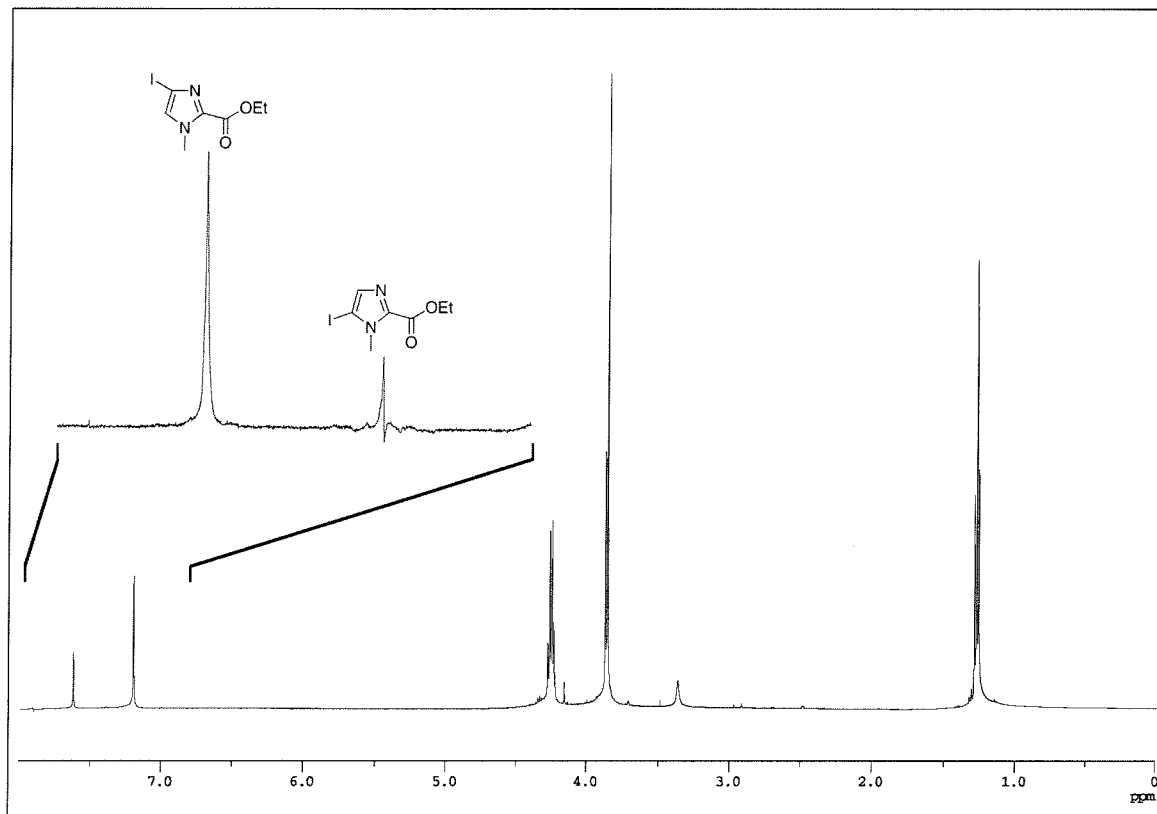


Figure 9. NOE difference spectrum of a mixture of **4** and **5** iodo imidazoles. The ¹H NMR spectra is shown full scale. The inset shows the NOE difference spectrum obtained after irradiation of the methyl peaks at 3.85 ppm. The spectra was obtained on a Bruker AM-500 NMR.

Coupling of the 4-iodo imidazole derivatives with acetyleno derivative **11** proceeded smoothly under Songashira conditions to give the alkynyl dimer **26**.^{12,13} Under the same conditions, the 4-bromo compound **18** failed to react even with an electron rich alkyne, phenylacetylene.²⁴ Yields for other palladium mediated coupling reactions on modified imidazoles have also been improved through the use of iodo rather than bromo derivatives.^{10,22}

This nitro alkyne was reduced by hydrogenation under H₂ with Pd on carbon as a catalyst. Subsequent Boc protection proved more difficult than anticipated. Reaction under a variety of conditions led to multiple products, several of which seemed to have demethylated the imidazole ring. Reduction in the presence of di-*t*-butyldicarbonate in

DMF in the absence of base afforded the polar product in moderate yield. Cleavage of the ester under basic conditions afforded the Boc diimidazole acid **28**.

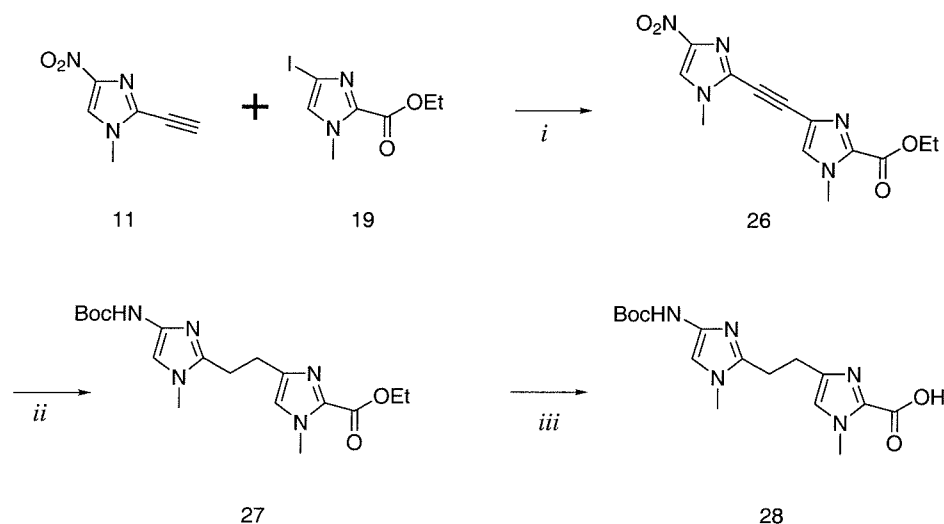


Figure 10. Final Sonagashira coupling to give alkyne 26. i) $\text{Pd}(\text{PPh}_3)_2\text{Cl}_2$, CuI, DMF, DIEA; ii) H_2 , 10 % Pd/C, Boc_2O ; iii) NaOH, MeOH/ H_2O .

Dimer **28** was activated with DCC/HOBt and coupled to a polyamide on the solid support to give polyamides **2** and **3**. These compounds were purified on a small scale, and their structures confirmed by MALDI-TOF mass spectrometry.

The original difficulty in reducing and Boc protecting the alkyne dimer led us to consider an alternate route. It was considered that the harsh conditions necessary to achieve the Boc protection might also lead to ring acylation and degradation. Early on, we had considered performing the entire sequence starting from Boc-Im-COOH. The Hunsdiecker-Borodin sequence on this starting material was less efficient than on the nitro so it was not actively pursued, but may ultimately be a useful route. As our initial efforts focused on the synthesis of only the $\text{ImIm}(\text{CH}_2)_2\text{ImIm}-\gamma\text{-PyPyPyPy}-\beta\text{-Dp}$ compounds, it seemed as if starting from the dimer, **32**, might be a plausible alternative route. Notably, dimer **32**, first synthesized by Ralf Jäger, is a useful intermediate for the solid phase synthesis of polyamides. Some progress has been made along this route, namely the

synthesis of compounds **33** and **35**. Initial coupling attempts gave only a small amount of what was putatively assigned as compound **36**. With some effort, this route should yield the desired trimer **38**. As the dimer route is nearing completion, and as a result of its more general applicability, the trimer route has not been pursued fully.

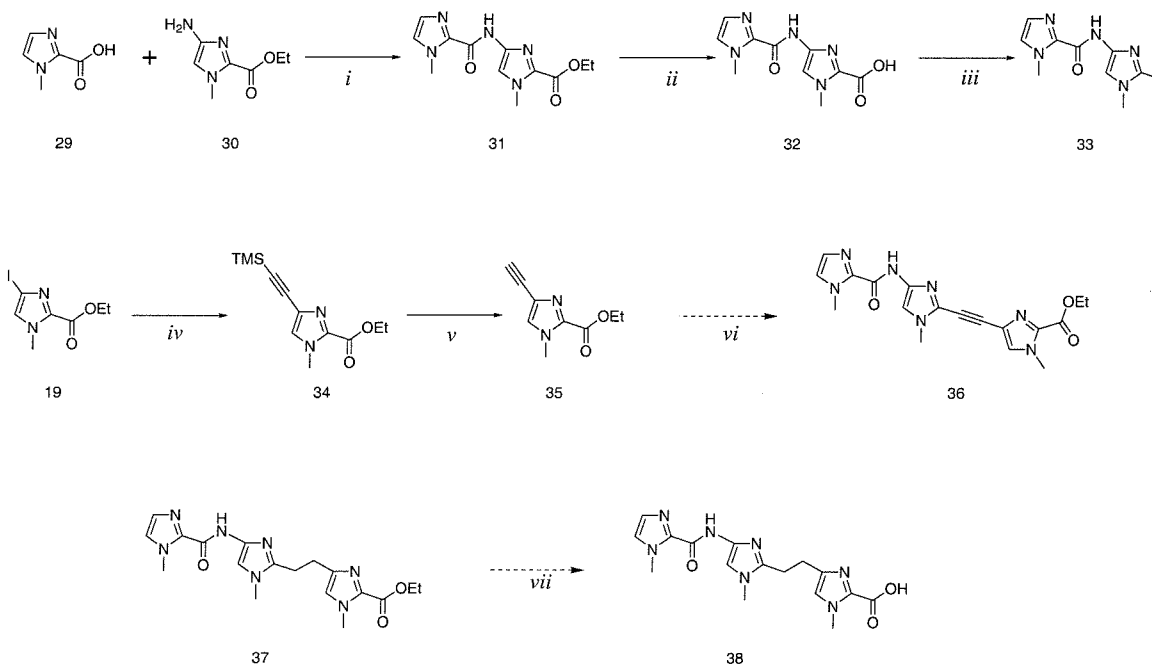


Figure 11. Alternative route to give trimeric ImImCH₂CH₂Im-COOH. *i*) HBTU, DIEA, DMF, 60%; *ii*) KOH, MeOH/H₂O, 95%; *iii*) NIS; *iv*) trimethylsilylacetylene, Pd(PPh₃)₂Cl₂, CuI, DIEA, DMF, 60 °C; *v*) NaOEt, THF; *vi*) **33**, Pd(PPh₃)₂Cl₂, CuI, DIEA, DMF; *vii*) KOH, MeOH/H₂O.

5.3 DNA Binding

The DNA binding properties of the alkyl linked polyamides were determined by quantitative DNase I footprint titrations. The results of the footprinting analysis are shown in table 1. Polyamide **2** specifically recognized a 6 base pair 5'-TGGGGA-3' binding site with an affinity of $9.0 \times 10^6 \text{ M}^{-1}$. While the compound did not display a marked increase in binding affinity over the parent hairpin polyamide **1**, it did retain specificity over the single base mismatch site 5'-TGGCGA-3'. As the compound binds non-specifically at higher concentrations, the binding constants of the mismatch sites can only be estimated as ...

Table 1. Apparent equilibrium association constants.^a

Polyamide	5'-TGGGGA-3'
ImImImIm- γ -PyPyPyPy- β -Dp ⁴ (1)	3.7×10^7
ImIm~ImIm- γ -PyPyPyPy- β -Dp (2)	9.0×10^6
ImIm~ImIm- γ -Py β PyPy- β -Dp (3)	$< 1 \times 10^6$

^a Binding constants were determined in collaboration with R. Bürli. Compounds were footprinted against a 3'-radiolabeled *PvuII/EcoRI* restriction fragment, pSEShp19.⁴ Compounds were equilibrated for 12 hours under the following solution conditions: 10mM Tris-HCl, 10mM KCl, 10mM MgCl₂, 5mM CaCl₂, and pH 7.0.

Replacement of an amide linkage by an alkyl linkage in a polyamide formally removes a single hydrogen bond in the bound structure. The slight energetic penalty suggests that either this hydrogen bond is not fully formed in the bound structure, or that some energetic gain was obtained through either increased flexibility or increased van der Waals interactions with the floor of the minor groove.

Furthermore, these results provide broader implications for the development of specific high affinity ligands targeting homo G-rich DNA sequences using the polyamide approach. It is possible that the sequence dependent conformation of such G-rich sequences is such that the minor groove is narrowed or less accessible. More likely, the contiguous row of 2-amino groups may prevent the polyamide from adopting an

energetically favorable complex within the groove. The polyamide may not sit as deeply in the groove when bound to such sequences, preventing the favorable van der Waals interactions between the ligand and the walls of the minor groove.

5.4 Conclusions

The synthesis of alkyl linked polyamides **2** and **3** for the recognition of G-rich DNA sequences has been reported. Key methodology for the synthesis is reported that lead to the completed synthesis of this interesting new family of compounds. Specifically, the carbon skeleton of the desired compound has been constructed in a regiospecific fashion through use of a Hunsdiecker-Borodin iodo decarboxylation Sonagashira coupling sequence.

The DNA binding data show that such modifications are tolerated within the motif but fail to increase the affinity of the hairpin molecules for DNA. The challenge of targeting homo G-rich DNA sequences remains. It is possible that such sequences adopt conformations that require significant alteration for polyamide binding. It is also possible that the contiguous 2-amino groups of the adjacent guanine residues form a steric blockade that prevents the polyamide from forming high affinity interactions with the DNA. Such interactions especially include the extensive van der Waals contacts formed when the polyamide sits deeply in the groove.

It is also worthy to note that the halogenated imidazole precursors could be useful for other polyamide studies. White, Baird, and Dervan proposed that orientational preferences of polyamide binding might correspond to dipole-dipole interactions with the DNA.²⁵ The strong electron withdrawing nature of the halo substituents could be used to test this hypothesis, particularly through substitution at the 5 position of the imidazole which should not sterically interfere with DNA binding. Secondly, it has been shown that

other halogenated heterocycles are useful as photochemical DNA cleavage agents via a proposed oxygen independent carbon radical formation.²⁶⁻²⁸

5.5 Experimental

All reagent grade chemicals were purchased from Aldich Chemical. Solvents were reagent grade from with the exception of DMF that was purchased as peptide synthesis grade from Perkin Elmer. All NMR spectra were collected on a 300 MHz GE-300 NMR. All spectra were referenced to residual solvent where possible, and to tetramethylsilane if the residual solvent peak was not observed. With the exception of the palladium couplings, all reactions were carried out under ambient atmosphere.

1-methyl 4-nitroimidazole (6). 1 g of 4-nitroimidazole (8.8 mmol) was refluxed in absolute ethanol the presence of 836 μ L dimethyl sulfate (8.8 mmol, 1eq.) for 48 hours. The reaction was quenched with 1M NaOH at reflux for 1 hour then extracted with dichloromethane. The compound was washed with 1M NaOH, and dried to give pure 1-methyl-4-nitroimidazole, 0.45g (3.5 mmol, 40%) ^1H NMR (CDCl_3) 3.95 (s, 3H), 7.53 (s, 1H), 7.92 (s, 1H). ^{13}C NMR δ 35.2, 133.0 141.5.

1-methyl 2-bromo 4-nitroimidazole (7). 10 g (58.4 mmol) of imidazole carboxylate 9 and 14.4 g (64.2 mmol, 1.1 eq.) of *N*-iodosuccinimide were heated to 80 $^\circ\text{C}$ in 150 mL of toluene and 5 mL of DMF for 30 minutes until evolution of CO_2 ceased. The reaction mixture was diluted with 200 mL of dichloromethane and washed with saturated sodium thiosulfate then brine. The organic phase was dried (Na_2SO_4) and evaporated, then flashed using 3-5% methanol in dichloromethane to give 8.9 g (43 mmol, 75%) of 7. ^1H NMR (CDCl_3) δ 3.75 (s, 3H), 7.81 (s, 1H).

1-methyl 2-trimethylsilylacetyleno 4-nitroimidazole (10). 1.3 g (7.6 mmol) of 9 and 1.7 g of *N*-iodosuccinimide were heated in 25 mL of toluene and 5 mL of DMF to

80 °C for 45 minutes. The reaction was partitioned between ethyl acetate and brine (200 mL each). The ethyl acetate layer was washed with saturated sodium thiosulfate and saturated NaHCO₃ then dried and evaporated to give crude 2 iodo compound **8**. This crude compound was dissolved in 20 mL of DMF and 10 mL of diisopropylethylamine. The solution was bubbled vigorously with Ar to degas. After 15 minutes, 100 mg of Pd(PPh₃)₂Cl₂ (0.14 mmol, 1.8 mol %) was added and the solution degassed for a further 15 minutes. Excess trimethyl silylacetylene (2 mL) was added, and the solution degassed further. Finally, 100 mg (0.53 mmol, 7 mol %) of CuI was added. The solution was slowly heated to 60 °C with continuous purging of Ar. The reaction turns from orange-red to black upon heating. After 1 hour, the reaction was partitioned between ethyl acetate and brine (100 mL each), then dried and evaporated. The product was purified by column chromatography using 2:1 hexanes: ethyl acetate to yield 307 mg (18% over 2 steps). ¹H NMR (DMSO-d₆) δ 0.27 (s, 9H), 3.74 (s, 3H), 8.49 (s, 1H).

1-methyl 2-acetyleno 4-nitroimidazole (11). 300 mg (1.3 mmol) of **10** were dissolved in 50 mL THF. 10 mL of an approximately 0.1M solution of sodium ethoxide was added. Reaction proceeded quickly to give the more polar product. Upon completion (5-20 minutes) the reaction was partitioned between brine and ethyl acetate. The ethyl acetate layer was dried and evaporated to give pure **11** in almost quantitative yield (200 mg). This material was used in its crude form for the synthesis of the alkyne dimer, **26**.

1-methyl 2-iodoimidazole (14). According to the published route,^{20,21} 12.3 g of *N*-methylimidazole (150 mmol) was cooled to -78 °C in 100 mL of Et₂O. 60 mL of 2.5M BuLi (150 mmol) in hexanes was slowly added to the solution. The solution was allowed to warm to -50 °C then cooled again to -78 °C. 38 g (150 mmol, 1 eq.) of I₂ was slowly added as a solution in 150 mL THF. The solution was allowed to warm to room

temperature and was extracted with 1M sodium thiosulfate. The organic phase was dried and evaporated to give 26.2 g (126 mmol, 84%) of product. ^1H NMR (CDCl_3) δ 3.59 (s, 3H), 7.00 (d, 1H, $J=1.3\text{Hz}$), 7.04 (s, 1H, $J=1.4\text{Hz}$) ^{13}C NMR (CDCl_3) 36.56, 90.65, 123.99, 132.40.

1-methyl 2-trimethylsilylacetylenoimidazole (15). 4.9 g (23.6 mmol) of **14** and 1.65 g (2.35 mmol, 0.1 eq.) $\text{Pd}(\text{PPh}_3)_2\text{Cl}_2$ and 0.67 g CuI were dissolved in 25 mL DMF and degassed by bubbling vigorously with Argon for 30 minutes. 5 mL of diisopropylethylamine was added and the solution degassed a further 5 minutes. 4.6 g (47 mmol, 2 eq.) of trimethylsilylacetylene was then added with continuous degassing. The solution was gradually heated to 80 °C over the course of an hour, then cooled. The solution was partitioned between brine and ethyl acetate (200 mL each). The organic phase was dried (Na_2SO_4) and evaporated. Chromatography on silica gel using 2:1 hexanes: ethyl acetate gave 1.7 g (9.4 mmol, 40%) of pure product **15**. ^1H NMR (CDCl_3) 0.249 (s, 9H), 3.70 (s, 3H), 6.86 (s, 1H), 7.00 (s, 1H).

ethyl 4(5)-bromo-1-methylimidazole 2-carboxylate (18, 20, 22). 2.0 g (13 mmol) of the ethyl 1-methylimidazole 2-carboxylate¹⁶ and 2.3 g *N*-bromosuccinimide were refluxed in THF overnight. The reaction was partitioned between brine and ethyl acetate (200 mL). The ethyl acetate layer was washed with saturated sodium thiosulfate until clear. The solution was dried (Na_2SO_4) and evaporated to give a mixture of mainly 4 and 5 bromo compounds. Careful (and sometimes repeated chromatography) in 2.5% MeOH in dichloromethane gave pure white compounds, **18**, **19**, and **20**. The product with the highest R_f value was assigned as the dibromo compound, **22**. The next compound to elute was the 4-bromo isomer, **18**. The final compound to elute was the 5-bromo isomer, **20**. Ethyl-4,5-dibromo 2-imidazole carboxylate, **22**: ^1H NMR (CDCl_3) δ 1.37 (s, 3H,

$J=6.9\text{Hz}$), 3.99 (s, 3H), 4.37 (q, 2H, $J=6.9\text{ Hz}$). Middle R_f : ethyl 4-bromo-1-methylimidazole carboxylate, **18**: ^1H NMR (CDCl_3) δ 1.29 (t, 3H, $J=6.9\text{ Hz}$), 3.95 (s, 3H), 4.35 (q, 2H, $J=7.2\text{Hz}$), 6.98 (s, 1H). ^{13}C NMR (CDCl_3) 14.13, 35.94, 61.60, 115.46, 125.52, 136.19, 158.13. Low R_f : ethyl 5-bromo-1-methyl-imidazole carboxylate, **20**: ^1H NMR (CDCl_3) δ 1.29 (t, 3H, $J=7.2\text{ Hz}$), 3.95 (s, 3H), 4.35 (q, 2H, $J=7.2\text{Hz}$), 7.11 (s, 1H).

ethyl 4(5)-iodo-1-methylimidazole 2-carboxylate (19 and 21). 15.4 g (0.1 mol) of the ethyl 1-methyl imidazole 2-carboxylate and 25.4 (0.1 mol) g I_2 were heated to $100\text{ }^\circ\text{C}$ in DMF for 5 days. The incomplete reaction was worked up by partitioning the reaction between brine and ethyl acetate (200 mL). The ethyl acetate layer was washed with saturated sodium thiosulfate until clear. The solution was dried (Na_2SO_4) and evaporated to give a mixture of 4 and 5 iodo compounds. Chromatography in 2.5% MeOH in dichloromethane gave pure white compounds 21 and 23. Spectral data for the 4-iodo isomer are identical with those obtained below. The next compound to elute was the 4-iodo isomer. The final compound to elute was the 5-iodo isomer. Ethyl 5-iodo -2-imidazole carboxylate (**21**) ^1H NMR (CDCl_3) δ 1.42 (s, 3H, $J=7.1\text{Hz}$), 4.01 (s, 3H), 4.40 (q, 2H, $J=7.1\text{Hz}$), 7.27 (s, 1H).

ethyl 4,5-diiodo-1-methylimidazole 2-carboxylate (23). From attempts to use more reactive iodination reagents, it was found that treatment of 17 with excess I_2/AgNO_3 at $100\text{ }^\circ\text{C}$ in DMF led to the efficient synthesis ethyl 4,5-diiodo-1-methylimidazole 2-carboxylate which could be isolated as faint orange crystals. ^1H NMR (CDCl_3) δ 1.36 (s, 3H, $J=7.1\text{Hz}$), 4.04 (s, 3H), 4.37 (q, 2H, $J=7.1\text{Hz}$).

ethyl 4-iodo-1-methylimidazole 2-carboxylate (19). 1-methyl-2-

trichloromethylketoimidazole (**24**) was synthesized according to Shibuya *et al.*¹⁹ on the gram scale or on kilogram scale with the modifications of N. Wurtz. 10 g (44 mmol) was dissolved in 50 mL of DMF and cooled to $-30\text{ }^{\circ}\text{C}$ with dry ice/acetonitrile. Separately, to a solution of iodine (8 g, 31.5 mmol) in 25 mL DMF was added a solution of 8 g (31 mmol) silver triflate in 25 mL of DMF. Upon addition a white precipitate, AgI, was formed. The mixture was slowly added to the imidazole over 15 minutes. The reaction was stirred and allowed to come to room temperature overnight. The reaction was filtered through Celite to remove the silver iodide. The solution was partitioned between brine and ethyl acetate, the ethyl acetate layer dried (Na_2SO_4) and evaporated. The product was taken up in absolute ethanol (100 mL) and ~ 0.5 g of sodium metal was added. After stirring for 1 hour, the reaction was partitioned between 0.1M HCl and ethyl acetate. The ethyl acetate layer was washed with saturated sodium thiosulfate, dried and evaporated to give crude **19**, that was purified by flash chromatography using 1.5% MeOH in CH_2Cl_2 to give pure white product, 3.1 g (11.1 mmol, 36%). ^1H NMR (CDCl_3) δ 1.42 (t, 3H, $J=7.1$ Hz), 4.00 (s, 3H), 4.42 (q, 2H, $J=7.1\text{Hz}$), 7.13 (s, 1H); ^1H NMR ($\text{DMSO}-d_6$) δ 1.28 (t, 3H, $J=7.1$ Hz), 3.86 (s, 3H), 4.26 (q, 2H, $J=7.1\text{Hz}$), 7.66 (s, 1H).

imidazole alkyne dimer (26). 1.6 g (7.2 mmol) of nitro trimethylsilylalkyne **10** was dissolved in 50 mL THF. 20 mL of $\sim 0.1\text{M}$ NaOEt was added, and the solution stirred until TLC shows complete conversion. (1:1 Hexanes:EtOAc starting material $R_f=0.9$; product $R_f=0.5$). The solution was partitioned between brine and ethyl acetate (200 mL each), then dried and evaporated. Separately, 2.42 g (8.6 mmol, 1.2 eq.) of 4-iodoimidazole **19** was dissolved in 10 mL of DMF and degassed with vigorous Ar bubbling for 20 minutes. Degassing was continued throughout. 400 mg of $\text{Pd}(\text{PPh}_3)_2\text{Cl}_2$ (0.57 mmol, 8 mol %) were added. A solution of the freshly prepared alkyne was added, followed by

excess (~5 mL) diisopropylethylamine. After a further 5 minutes of degassing, 200 mg (1.0 mmol, 13 mol %) of CuI were added. The reaction was slowly heated to 60 °C. After 1 hour, the reaction was partitioned between ethyl acetate and brine. The ethyl acetate was dried and evaporated, and the product flashed using 2.5% methanol in dichloromethane to yield 748 mg (2.5 mmol, 34%). ¹H NMR (DMSO-*d*₆) δ 1.31 (t, 3H, *J*=7.2Hz), 3.81 (s, 3H), 3.94 (s, 3H), 4.31 (q, 2H, *J*=7.1Hz), 8.07 (s, 1H), 8.55 (s, 1H). ¹H NMR (CDCl₃) δ 1.43 (t, 3H, *J*=7.1 Hz), 3.86 (s, 3H), 4.05 (s, 3H), 4.44 (q, 2H *J*=7.1 Hz), 7.39 (s, 1H), 7.77 (s, 1H). ¹³C NMR (CDCl₃) δ 14.5, 35.7, 36.2, 38.8 61.7, 115.0, 126.2, 128.6, 132.1, 136.9, 138.5, 156.6, 159.1.

Boc imidazole alkane dimer ester (27). A solution of 5 mg of alkyne dimer **26** was stirred in 1 mL of DMF with 5 mg of 10% Pd on carbon in the presence of excess (20 mg) of di-*t*-butyldicarbonate for 48 hours under 400 psi (27 atm) of H₂ in a stainless steel pressure vessel (Parr Instruments). The reaction mixture was diluted with dichloromethane, filtered through celite and evaporated. The crude product was purified by silica gel chromatography using 5% methanol in dichloromethane. ¹H NMR (CDCl₃) δ 1.41 (t, 3H), 1.48 (s, 9H), 3.49 (s, 3H), 1.8-3.05 (m, 4H), 3.92 (s, 3H), 4.40 (q, 2H), 6.84 (s, 1H), 7.43 (s, 1H). Electrospray MS obs. (M+H)⁺ 378.4 calcd. 378.2.

Boc imidazole alkane dimer acid (28). A solution of 120 mg of alkyne dimer **26** was stirred in 1 mL of MeOH with 1 mL of 1M KOH. After 2 hours at room temperature, the pH was adjusted to between 3 and 4. The solution was extracted with ethyl acetate. The aqueous layer was neutralized and lyophilized to afford a white powder. Titration of the powder with methanol, and evaporation of the methanolic solution, gave **27** as a white solid contaminated with a large quantity of salt. The crude product was used for solid phase polyamide synthesis without further purification. ¹H NMR (CD₃OD) δ 1.53 (s, 9H),

3.16 (t, 2H, $J=7.2$ Hz), 3.36 (t, 2H, $J=7.6$ Hz), 3.84 (s, 3H), 4.13 (s, 3H), 7.16 (s, 1H), 7.44 (s, 1H). Electrospray MS obs. (M+H)⁺ 348.2 calcd. 350.2.

diimidazole acid, 32. Imidazole acid **29**, 5.0 g (40 mmol), was activated at 60 °C for 1 hour with 14.3 g (38 mmol) HBTU in 70 mL DMF with 20 mL DIEA. 8.15 g (40 mmol) of the hydrochloride salt of amine **30** was added to the activated acid, and the solution stirred for 12 hours at 60 °C. The solution was precipitated into 200 mL of ice water and filtered to give crude **31**. A small sample was purified by flash chromatography with 5% MeOH/CH₂Cl₂ to give analytically pure **31**. ¹H NMR (CDCl₃) δ 1.42 (t, 3H, $J=7.1$ Hz), 4.00 (s, 3H), 4.05 (s, 3H), 4.40 (q, 2H, $J=7.1$ Hz), 6.97 (s, 1H), 7.02 (s, 1H), 7.54 (s, 1H), 9.69 (s, 1H). The crude ester was heated at 60 °C in 200 mL of 1M KOH for 4 hours, then acidified to give 5.5 g (22 mmol, 55%) of dimeric acid **32**. ¹H NMR (DMSO-*d*₆) δ 4.05 (s, 3H), 4.11 (s, 3H), 7.31 (s, 1H), 7.63 (s, 1H), 7.77 (s, 1H).

ethyl 1-methyl-4-trimethylsilylacetylenoimidazole 2-carboxylate, 34. 1.0 g (3.6 mmol) of 4-iodo imidazole **19** was dissolved in 20 mL of DMF with 5 mL of DIEA. The solution was degassed with Ar for 15 minutes after which 100 mg of Pd(PPh₃)₂Cl₂ (0.14 mmol, 3.8 mol %) was added. After 5 additional minutes, excess trimethylsilylacetylene (2 mL) was added. The solution was degassed a further 5 minutes before the addition of 100 mg of CuI. The reaction was heated to 60 °C during which time it turned from orange to black. The solution was heated at 60 °C for one hour then partitioned between 100 mL of ethyl acetate and brine. The organic phase was dried and evaporated to give the crude product. The residue was dissolved in dichloromethane and loaded on a silica gel column equilibrated in 2:1 hexanes: ethyl acetate. The product was eluted with 2:1 hexanes: ethyl acetate. 433 mg (1.7 mmol, 48%). ¹H NMR (CDCl₃) δ 0.12 (s, 9H), 1.31 (t, 3H), 3.89 (s, 3H), 4.28 (q, 2H, $J=7.2$ Hz), 7.11 (s, 1H). ¹³C NMR (CDCl₃) δ 14.0, 35.8, 61.4, 945.0, 96.9, 124.1, 129.5, 136.0, 158.4.

ImIm~ImIm- γ -PybPyPy- β -Dp (3). H₂N-Im- γ -Py β PyPy- β -PAM was synthesized starting from 1.0 g of β -alanine PAM resin (Peptides International, KY) 0.25 mmol/g substitution by standard procedures.¹⁸ Boc dimer acid 28 was activated with dicyclohexylcarbodiimide and hydroxbenzotriazole in DMF and added to the amino resin with diisopropylethylamine. The reaction vessel was shaken for 3 hours at 32 °C then at room temperature for 12 hours. The Boc group was removed with 100% trifluoroacetic acid (2 x 10 minutes), and the polyamide capped with 1-methylimidazole-2-carboxylate as described.¹⁸ Cleavage from the resin with *N,N*-dimethylaminopropylamine for 12 hours at 55 °C afforded crude **3** that was then purified by reversed phase HPLC. MALDI-TOF (M+H)⁺ obs. 1158.61 calcd. 1158.57.

5.6 References

- (1) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380.
- (2) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8385.
- (3) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.
- (4) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953-6961.
- (5) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198-8206.
- (6) Coy, D. H.; Hocart, S. J.; Sasaki, Y. *Tet.* **1988**, *44*, 835-841.
- (7) Sasaki, Y.; Coy, D. H. *Peptides* **1987**, *8*, 119-121.
- (8) Yamamoto, Y.; Kimachi, T.; Kanaoka, Y.; Kato, S.; Bessho, K.; Matsumoto, T.; Kusakabe, T.; Sugiura, Y. *Tet. Lett.* **1996**, *37*, 7801-7804.
- (9) Heck, R. F. *Org. React.* **1982**, *27*, 345-389.
- (10) Shapiro, G.; Gomez-Lor, B. *J. Org. Chem.* **1994**, *59*, 5524-5526.
- (11) Abarca-Gonzalez, B.; Jones, R. A.; Medio-Simon, M.; Quilez-Pardo, J.; Sepulveda-Arques, J.; Zaballos-Garcia, E. *Synth. Comm.* **1990**, *20*, 321-331.
- (12) Stephens, R. D.; Castro, C. E. *J. Org. Chem.* **1963**, 3313-3315.
- (13) Sonagashira, K.; Tohda, Y.; Hagihara, N. *Tet. Lett.* **1975**, *50*, 4467-4470.
- (14) Benjes, P.; Grimmett, R. *Heterocycles* **1994**, *37*, 735-738.
- (15) Suwinski, J.; Salwinska, E.; Watras, J.; Widel, M. *Pol. J. Chem.* **1982**, *56*, 1261-1272.
- (16) Krowicki, K.; Lown, J. W. *J. Org. Chem.* **1987**, *52*, 3493-3501.

- (17) Naskar, D.; Chowdhury, S.; Roy, S. *Tet. Lett.* **1998**, 39, 699-702.
- (18) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6141-6146.
- (19) Nishiwaki, E.; Tanaka, S.; Lee, H.; Shibuya, M. *Heterocycles* **1988**, 27, 1945-1952.
- (20) Tian, Z.-Q. *Mechanisms of iron porphyrin catalyzed epoxidation of alkenes and the concomitant N-alkylhemin formation*; University of California: San Diego, 1992.
- (21) Traylor, T. G.; Hill, K. W.; Tian, Z.-Q. *J. Am. Chem. Soc.* **1988**, 110, 5571-5573.
- (22) Phillips, J. G.; Fadnis, L.; Williams, D. R. *Tet. Lett.* **1997**, 38, 7835-7838.
- (23) Shinkai, I. *Five-membered Rings with two heteroatoms and fused carbocyclic derivatives*; 1st ed.; Pergamon: Oxford; New York, 1996; Vol. 3.
- (24) Sakamoto, T.; Shiga, F.; Yasuhara, A.; Uchiyama, D.; Kondo, Y.; Yamanaka, H. *Synth. Stuttgart* **1992**, 746-748.
- (25) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, 119, 8756-8765.
- (26) Sakai, Y.; Matsumoto, T.; Tanaka, A.; Shibuya, M. *Heterocycles* **1993**, 36, 565-573.
- (27) Matsumoto, T.; Utsumi, Y.; Sakai, Y.; Toyooka, K.; Shibuya, M. *Heterocycles* **1992**, 34, 1697-1702.
- (28) Nishiwaki, E.; Nakagawa, H.; Takasaki, M.; Matsumoto, T.; Sakurai, H.; Shibuya, M. *Heterocycles* **1990**, 31, 1763-1767.

Chapter 6:

Studies on the *isoG-isoC* Base Pair

6.1 Abstract

The non-natural base pair between isoguanosine and isocytosine has been studied as a means of enzymatically incorporating a modifiable nucleoside into RNA for autocleavage studies. We have enzymatically incorporated the functionalizable nucleoside *N*⁶-(6-Aminohexyl)-isoguanosine into a hammerhead ribozyme using T7 RNA polymerase by the method of Tor and Dervan.¹ Post-transcriptional modification of the ribozyme with Fe(II)-EDTA yielded a species competent for oxidative cleavage; however, no cleavage distal to the site of incorporation was observed in the hammerhead system. The enzymatic recognition the base pair between d-isoguanosine (d-isoG) and 5-methylisocytosine triphosphate (Me_{iso}CTP) has been probed in the interest of site-specifically incorporating a non-natural intramolecular base pair into RNA biopolymers. d-isoG does not direct the incorporation of Me_{iso}C in transcription experiments utilizing T7 RNA polymerase. Instead, the enzyme incorporates uridine opposite d-isoG, suggesting the presence of a different tautomeric form of d-isoG.

6.2 Background

The field of RNA biochemistry has expanded quickly following the discovery of RNA catalysis in the early 1980's.^{2,3} In the traditional view of molecular biology, RNA was seen mainly as a messenger, directing protein synthesis, and as an adapter molecule for the decoding of the mRNA message. The discoveries of the roles of RNA in

regulation⁴ and catalysis^{5,6} have strengthened the emerging view of an RNA world, in which RNA functioned both genetically and catalytically. The functions of RNA biopolymers are clearly linked to their structural properties. The discoveries of the importance of RNA in self-splicing introns,⁵ in mRNA splicing,⁷ in ribosome catalysis,⁸ and in artificially evolved ribozymes⁹ have focused attention on the determination of RNA structures. Additionally, the identification of RNA structures important in disease, such as the TAR and RRE RNAs of HIV, marks RNA structures as potential therapeutic targets.¹⁰ The rapidly growing field of antisense research must also contend with the formation of various RNA structures.¹¹ Thus, a detailed understanding of the biochemistry of RNA is dependent on the determination of RNA structures.

As a result of the growing interest in RNA structure, physical methods for the determination of RNA structures are being developed. In spite of early success with the determination of the crystal structure of tRNA^{phe} in the 1970's,¹² RNA molecules have proven difficult to crystallize. The structure of an RNA duplex has been determined by crystallography,¹³ and the structure of tRNAs have continued to be studied,¹⁴ but few other structures have been solved by this method. More recently, structures of the hammerhead ribozyme¹⁵ and of the *Tetrahymena* group I intron have been solved¹⁶. The work described in this chapter precedes both structures.

As the progress of x-ray crystallography in this area has been slow, solution methods, mainly NMR, have become popular.¹⁷ The severe overlap of the proton signals limits the ability to study large RNA biopolymers by NMR without the labeling of atoms for heteronuclear experiments.¹⁸ Although progress had been reported with both x-ray crystallographic and solution NMR methods,¹⁹ these techniques faced the uncertainty of the stability of RNA tertiary structures.²⁰ Unlike proteins, it is possible that large RNA structure possess far greater conformational flexibility and will prove intrinsically difficult to probe.

As these high resolution techniques are still being refined, chemical and enzymatic probes have been widely used to probe the secondary and tertiary interactions of RNA.²¹ The prediction of the secondary structure of RNA is relatively reliable using computational and phylogenetic approaches.²² Computational methods for the prediction of RNA tertiary structures from the predicted secondary structures are, however, still in developmental stages and cannot provide reasonable *de novo* predictions of tertiary structure;²³ computational modeling has, however, led to reasonable descriptions for well characterized systems, such as the group I intron. RNA degrading enzymes are known that possess specificity for single and double stranded RNA. Chemical methods for the mapping of RNA structures include MPE-Fe(II) which is specific for base paired regions,²⁵ the non-specific cleavage reagent, Fe-EDTA,²⁶ and recently developed transition metal complexes, which are generally reactive at sites of tertiary interaction.²⁷ Additionally, chemical footprinting reagents which react with specific base pairs can be used to map protection of nucleoside functional groups.²¹ While these methods are extremely useful for the confirmation of secondary structure predictions, they have not proven as useful for the prediction of tertiary interactions. The use of affinity cleavage has also been successful using the reagent G-DTPA to cleave the tetrahymena ribozyme.²⁸ Most recently, the introduction of a modified uridine with a tethered EDTA complex into tRNA^{phe} has demonstrated the general possibility of autocleavage mapping of RNA structures.²⁹

We chose to investigate an extension of the work of Tor; Dervan using an enzymatic approach to the *incorporation of single modified nucleosides for affinity cleavage*.¹ Site-specific modification of oligonucleotides has been accomplished routinely through chemical synthesis³⁰ and more recently by Han and Dervan.²⁹ Additionally, T7 RNA polymerase has been utilized to enzymatically incorporate modified natural nucleosides opposite their natural complements, limiting the usefulness to sequences with unique sites, or to random incorporation throughout the sequence.³¹ This technique has been successfully combined with a ligation step to generate large ribooligonucleotides with

sequence specific modifications.³² The approach taken by Tor and Dervan is the first reported method for the enzymatic incorporation of a single modifiable nucleoside in a mixed RNA sequence. The specificity of T7 polymerase and the genetic code, in general, arise from the unique hydrogen bonding array presented by each nucleoside pair. The non-natural

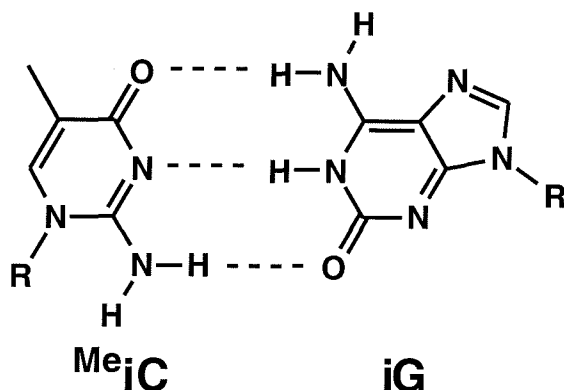


Figure 1. The base pair between ^{Me}isocytosine and isoguanosine.

base pair of 5-methylisocytosine with isoguanosine (figure 1) displays an array of hydrogen bond donors and acceptors recognized by A. Rich in 1962 to be distinct from the natural base pairs.³³ Benner and coworkers have shown that d-isoC in a DNA template can direct the incorporation of isoGTP into RNA, although with low efficiency.³⁴ It was also noted by these workers that isoC undergoes a fairly rapid deamination to form uracil. The technology was, however, used successfully to expand the genetic code to form a 65th codon for use in the incorporation of non-natural amino acids.³⁵ Tor and Dervan showed that DNA templates containing ^{Me}isoC could be transcribed to yield products with isoG or the modifiable nucleoside, *N*⁶-(6-Aminohexyl)-isoguanosine, at the appropriate position in the transcript in good yield.¹

The introduction of EDTA site-specifically into an RNA provides tertiary structural information through cleavage of positions proximal to the Fe(II)-EDTA moiety in three dimensions. The cleaved positions can then be mapped by high resolution electrophoresis, as demonstrated for a known structure.²⁹ We hoped to achieve the same end through enzymatic rather than chemical incorporation of derivatizable nucleosides. As a control system for this enzymatic technique, we chose to incorporate the modifiable nucleoside *N*⁶-

(6-Aminohexyl)-isoguanosine into a variant of the hammerhead ribozyme.³⁶ The potential ability to use this enzymatic system to *incorporate a non-natural, modifiable base pair into RNA* was explored in later experiments. Specifically, the ability of a DNA template containing a single d-isoG nucleotide to direct the incorporation of Me_{iso}CTP was studied. Lastly, results in these and other studies suggest that isoguanosine has interesting tautomeric properties. *Structural studies on the non-natural base pair of d-Me_{iso}C with d-isoG* and of the base pair of *isoG with T* were undertaken to help resolve this tautomeric ambiguity.

6.3 Application of the d-Me_{iso}C-AH-isoG system to the study of the Hammerhead Ribozyme

The hammerhead ribozyme was chosen as a model system for the application of the d-Me_{iso}C-AH-isoG system, as it was small, had an intrinsic assay for activity, and was well characterized catalytically. Additionally, the hammerhead is just out of the range of current NMR techniques. At the time of this work, its structure had not yet been determined. Hammerhead ribozymes are a conserved family of ribozymes derived from plant viroids.³⁶ The ribozyme has the general structure shown in figure 2.³⁷ Cleavage

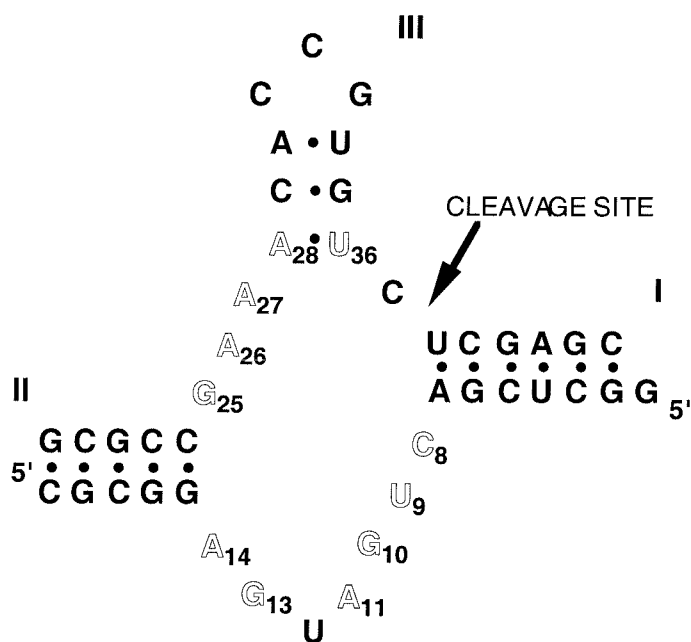


Figure 2. Consensus structure of the hammerhead ribozyme. Outlined bases are conserved positions. (After Ruffner *et al.*, 1990; Fu and McLaughlin, 1992)

takes place by a hydrolytic mechanism, involving attack of the 2' hydroxyl of C37, generating 2',3' cyclic phosphate and 5' hydroxyl termini. The presence of Mg²⁺ is required for catalysis, and was shown to be intimately involved in the active site, potentially through the interaction of a Mg-OH species.³⁸ Relatively small Mg²⁺-dependent NMR spectral shifts were observed, suggesting that the structure is preorganized to

bind magnesium and that magnesium plays a predominantly catalytic role.³⁹ The kinetics of hammerhead ribozyme cleavage reactions have been thoroughly studied by Uhlenbeck and coworkers.⁴⁰

The small size of the hammerhead ribozyme has made it distinctly suitable for atomic mutagenesis studies in which single functional groups or atoms are deleted from specific nucleosides. McLaughlin has pioneered this approach, identifying specific N7 nitrogens necessary for activity, as well as specific purine exocyclic functionalities.⁴¹ Most recently, Eckstein and coworkers have studied the importance of the guanosine residues within the central core, with replacements to inosine, nebularine, and isoguanosine.⁴² In spite of all of the specific data regarding the atomic involvement necessary for catalysis, a refined model of the hammerhead tertiary structure did not exist in the literature at the time of this work.

The site of modification chosen was the U12 position which is the only mutable nucleoside in the conserved CUGA loop of the ribozyme.³⁷ Additionally, a site specific crosslink from U12 to the nucleoside 5' to the cleavage site has been observed, suggesting that the two positions might be proximal in space.⁴³ The system chosen for this study was that used by Ruffner *et al.* which is extremely sensitive to perturbations in the catalytic core.³⁷ Kinetic assays were used, therefore, to relate the effect of modifications of the ribozyme at this position to structural perturbations in the core of the hammerhead ribozyme which effect the rate of catalysis. The kinetic measurements were made with excess ribozyme at a temperature well below the melting point of the structure, such that the kinetic measurements reflect single-turnover kinetics. The relative rates observed, therefore, reflect the rate of the catalytic step of the reaction.

Synthesis of d-Me_{iso}C phosphoramidite and AH-isoGTP were carried out as previously described.¹ DNA templates were prepared corresponding to the ribozyme strand of the hammerhead, with A, T, or Me_{iso}C at the position corresponding to position 12 in the ribozyme. These templates were transcribed and the RNA transcripts purified to yield hammerhead mutants with U12 substituted to A, isoG, and AH-isoG. The efficiency of incorporation of the modified nucleoside was measured indirectly by reaction of a radioactively labeled ribozyme RNA with the NHS-ester of Biotin. Gel shift analysis,

shows that the modification reaction shifts >95% of the population, indicating essentially no misincorporation of adenosine opposite the d-Me_{iso}C, in agreement with the results obtained by enzymatic digestion of the RNA transcripts.¹ AH-isoG containing hammerheads were post-transcriptionally modified with either the NHS-ester of biotin or with EDTA dianhydride. Post-transcriptionally modified hammerheads were repurified by gel electrophoresis such that the modification was present for all molecules.

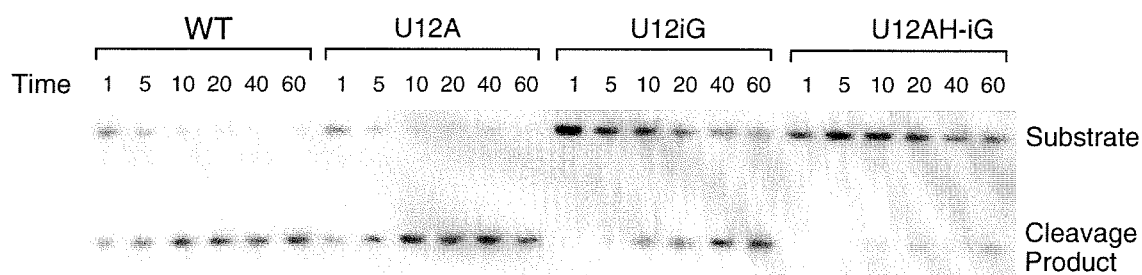


Figure 3. Data of a single kinetic experiment. A time course is shown for each of the unmodified mutants. Time course is shown in minutes. Experiments were performed as described in the experimental section.

The catalytic efficiency of the modified hammerheads was assayed as described.³⁷ Cleavage reactions to test the catalytic efficiency of the modified ribozymes yielded the relative rates shown in table 1. The relative rates observed for the introduction of the U12A mutation are consistent with the literature and serve as an internal control for the experimental system.³⁷ The introduction of the modified base isoG has a slight effect, though within the range for natural nucleosides. The AH-isoG lowered activity still, but still resulted in an active ribozyme.

Table 1. Relative rates of modified hammerhead ribozymes.

Ribozyme	k_{rel}
Wild type	1
U12A	1 ± 0.25
U12iG	0.28 ± 0.06
U12AHiG	0.09 ± 0.03
U12AHiG-EDTA	<0.005
U12AHiG-Biotin	<0.001

The results clearly indicate that the hammerheads modified with AH-isoG-EDTA or AH-isoG-Biotin have negligible activity. Thus, further modification of AH-isoG by attachment of molecules to the primary amino group appears to interfere with the formation of an active hammerhead structure.

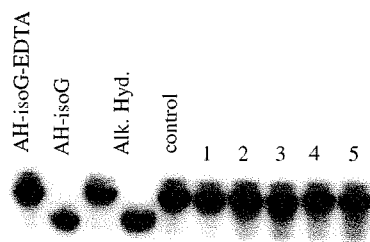


Figure 4. Attempted AH-iG-EDTA autocleavage of the U12-AH-isoG-EDTA hammerhead ribozyme. The 5' end of the ribozyme strand of the hammerhead is labeled in these experiments. Experiments were performed as described in the experimental section.

Even in the absence of catalysis, it was decided to attempt to map the pattern of EDTA-Fe(II) cleavage on the hammerhead ribozyme with the U12AH-isoG-EDTA. For these experiments, the EDTA containing ribozyme was annealed to a chemically synthesized substrate containing a single deoxynucleoside at the position 5' to the cleavage site. Lacking an internal nucleophile, these substrates are incompetent for cleavage. No EDTA-Fe(II) induced cleavage was observed either on the proximal (ribozyme) or on the distal (substrate) strand of the complex, under a wide range of conditions, including varied

iron concentrations and varied magnesium concentrations, figure 4, and other data not shown. As direct evidence for the presence of the EDTA moiety was not available, the AH-isoG-EDTA-containing RNA was subjected to autocatalysis while annealed to complementary DNA strand. Cleavage resulted,

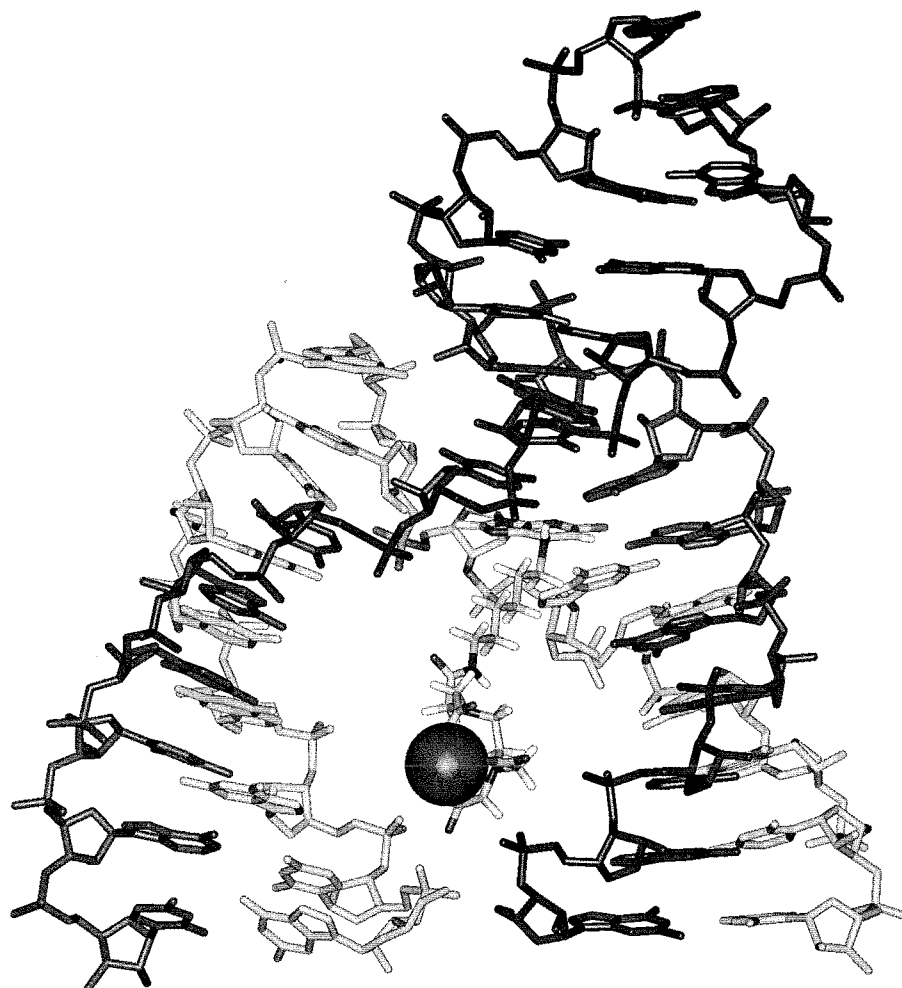


Figure 5. Model of hammerhead ribozyme containing AH-isoG-EDTA-Fe.

as predicted, in a manner consistent with the location of the EDTA moiety in the major groove of the double helical RNA-DNA hybrid.⁴⁴ Additionally, the cleavage pattern is consistent with that derived for RNA-DNA heteroduplexes.⁴⁵ The cleavage efficiency observed in these reactions was approximately 1%, low compared to cleavage efficiencies observed for heteroduplexes containing the EDTA modified nucleoside U*.⁴⁵

In order to provide a shorter linker to bring the EDTA closer to the RNA structure, and a test of T7 RNA polymerase's ability to accept a shorter linker, *N*⁶-(6-Aminopropyl)-isoguanosine triphosphate was synthesized, essentially by the same procedure as for AH-isoGTP (figure 6). In small scale transcription experiments (data not shown), the triphosphate was accepted by the enzyme comparably to the AH-isoGTP. Studies on the incorporation of AP-iGTP into the hammerhead ribozyme were not performed.

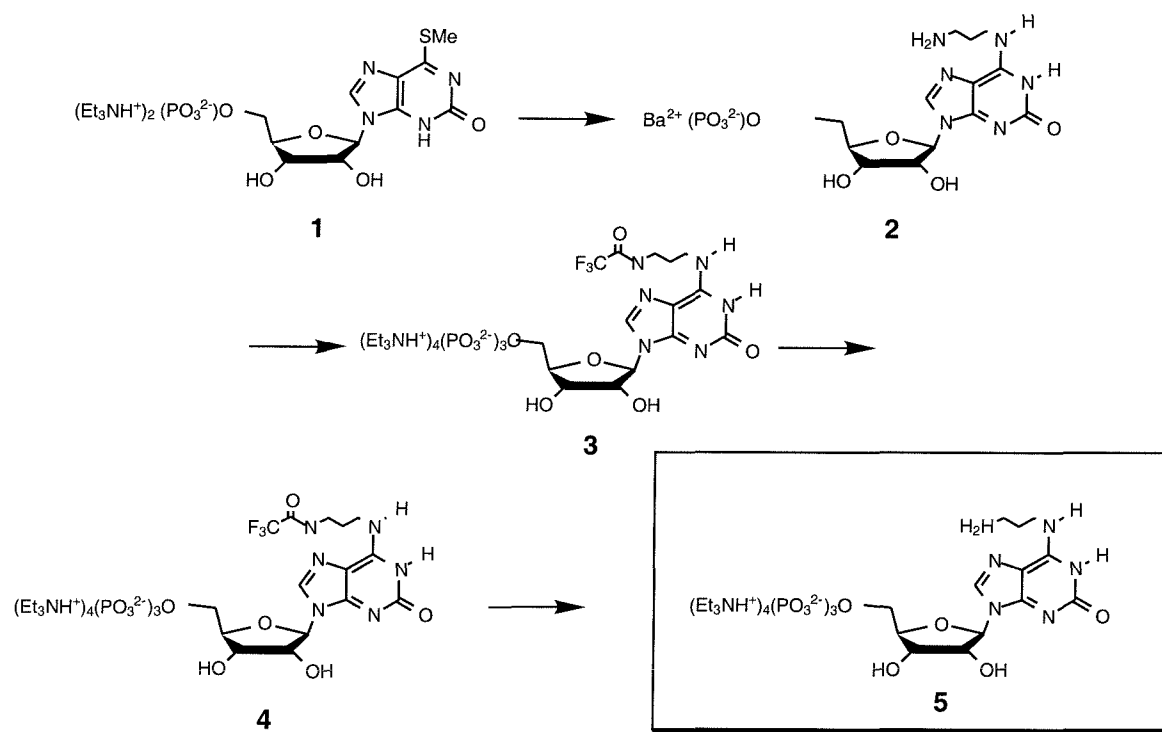


Figure 6. The synthesis of AP-iGTP, by the procedure of Tor and Dervan, 1993.

Conclusions. The introduction of the modified base AH-isoG-EDTA into the hammerhead ribozyme did not provide structural information regarding the tertiary structure of this catalytic RNA. It is possible that modifications to the catalytic core by any mechanism will not be tolerated in the hammerhead ribozyme. It will be important to realize in future studies that *RNA tertiary structures containing rationally selected modification sites may still be intolerant of such large perturbations*. The severity of the inhibition of catalysis further suggests that sites in large RNA's should be chosen some distance from the catalytic centers. The insertion of isoG, as a G or A analog, is still quite

feasible by this method. The insertion of isoG into hammerhead ribozymes, for example, required the use of a synthetic/enzymatic approach as the RNA phosphoramidite developed could only be incorporated at the 5' end of oligonucleotides.⁴² In the absence of a ribo-isoG phosphoramidite, this enzymatic method will still be useful for the incorporation of unmodified isoguanosine. The fidelity of transcription as measured above for the AH-isoG shows that RNA's can be modified almost quantitatively. The lowered transcription efficiencies involved in the incorporation of novel nucleosides (typically 10-50%, depending on sequence) by transcription indicate that this technique is best suited for the analysis of single RNA strands which can be radiolabeled. Difficulties in obtaining large quantities of modified, repurified unlabeled material hampered efforts to place the isoG in other positions of the hammerhead ribozyme. The lack of success in generating distal cleavage patterns, both by me and by Dr. Tor, indicate that there may be serious limitations to the use of modified isoGs in cleavage applications. Orientation of the nucleoside and linker arm away from the DNA in an extended conformation could lower the cleavage efficiency in the hammerhead complex to levels comparable to background cleavage. A similar lack of cleavage was observed in autocleavage studies of tRNA containing AH-isoG-EDTA.⁴⁶ The RNA-DNA heteroduplex results showed relatively inefficient cleavage. The success of Han and Dervan in employing a modifiable U phosphoramidite suggests, however, that the site-specific incorporation of an EDTA moiety by chemical means will be a useful tool for mapping of the features of RNA structures.²⁹ It is suggested that the use of chemical means for the incorporation of modifiable nucleosides into RNA will be a more profitable route for the study of RNA tertiary structure.

6.4 Studies on the Transcription of Templates Containing Deoxy-Isoguanosine

With the current ability to chemically introduce modifiable nucleosides into RNA oligomers, the potential advantages of a transcriptional system are *the ability to make long RNA oligomers*, and potentially, to *enzymatically introduce a non-natural base pair into*

RNA. Recently, Benner and coworkers published the results of a study of the incorporation of isoC into RNA and DNA by various enzymes.³⁴ Using T7 RNA polymerase, no full length transcription products were observed for templates containing isoguanosine in the

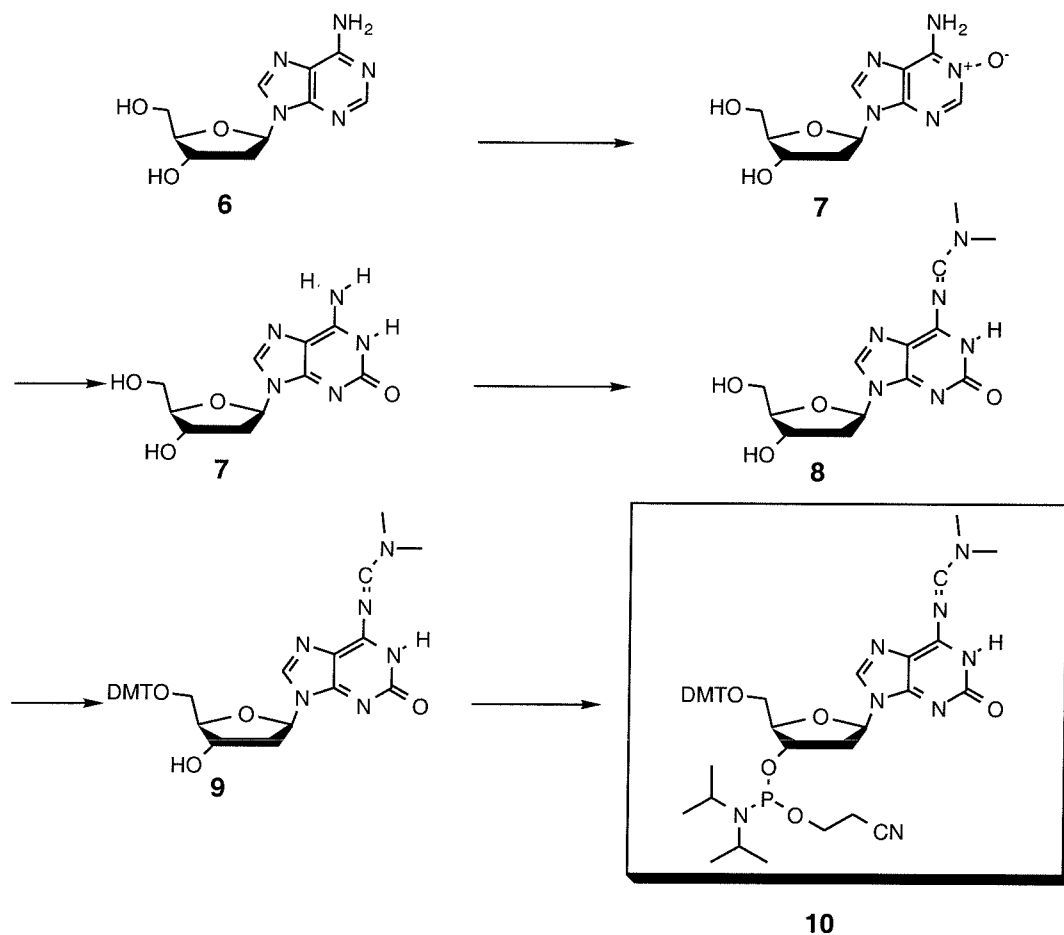


Figure 7. Synthesis of protected deoxyisoguanosine phosphoramidite, **10**. Details are given in the experimental section.

presence of isocytosine triphosphate and the absence of uridine triphosphate. Uridine was, however, efficiently incorporated into the RNA transcript in the presence or absence of isoCTP. The possibility existed that the failure of incorporation of isoC resulted from an *enzymatic preference for the natural nucleotide triphosphate*, UTP, over the non-natural isoCTP, or from an *inhibitory action of either isoCTP or an impurity* in the isoCTP

preparation on the polymerase. Consequently, it was hoped that substitution of Me_{iso}CTP for isoCTP, while providing a substrate less prone to deamination, might also interact more favorably with the enzyme.

Testing the transcriptional system, with templates containing d-isoG, required the efficient synthesis of the appropriately protected phosphoramidite of the nucleoside. Our interest in structural studies indicated that the published synthesis of Seela *et al.* was not feasible for the large scale synthesis of oligonucleotides containing d-isoG as it had an overall yield of <1%.⁴⁷ Alternatively, synthesis of d-isoG can be accomplished using the photochemical rearrangement described for the formation of isoguanosine.⁴⁸ In our hands the synthesis of compound **7**, deoxyisoguanosine, through this chemistry was reasonably efficient, and eliminated the necessity of protection of the sugar hydroxyls. A similar approach was employed on a protected rather than unprotected adenosine.³⁴ Isoguanosine was suitably protected for automated synthesis as previously described.⁴⁷ Phosphitylation yields were initially quite low, resulting from some phosphitylation at the O2 position. Treatment with an excess of phosphitylating agent followed by silica flash chromatography yielded the product **10** in moderate yield. This procedure gave an overall yield of 10%, approximately 10-fold higher than previously reported routes. Since this work was completed, a synthesis more efficiently protected d-isoG phosphoramidite has been reported.⁴⁹

Synthesis of the second component of the purification was performed using standard nucleoside chemistry. Ribothymidine (**11**) was synthesized by a known procedure.⁵⁰ The anhydro compound **12** was formed in good yield by a Mitsunobu reaction as described for uridine.⁵¹ Ammonolysis of **12** yielded the nucleoside 5-methylisocytosine (**13**) as reported for the formation of isocytosine.⁵¹ The triphosphate was formed directly from the nucleoside in moderate yield.⁵² Purification by ion exchange chromatography yielded pure 5-methylisocytosine triphosphate (**14**).

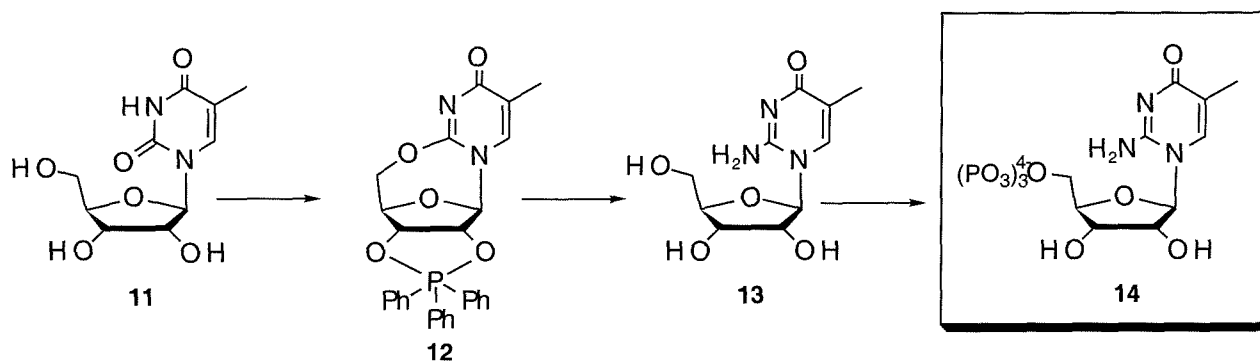


Figure 8. Synthesis of $\text{Me}_{\text{iso}}\text{CTP}$. Details are in the experimental section. Ph = phenyl.

The transcription system for testing the incorporation of $\text{Me}_{\text{iso}}\text{CTP}$ opposite d-isoG is shown in figure 10. Original experiments utilized $\alpha\text{-}^{32}\text{P}\text{-ATP}$ to end-label the transcripts. Later experiments were performed with $\alpha\text{-}^{32}\text{P}\text{-GTP}$ in order to body-label the transcripts. The phosphorimaged gel of the transcription experiment is shown in figure 10. Lane 2 shows the normal transcription of a natural template in the presence of $\text{Me}_{\text{iso}}\text{CTP}$, with no inhibition compared to the control (lane 1, no $\text{Me}_{\text{iso}}\text{CTP}$). Only in the presence of UTP were full length transcripts synthesized (lanes 3 and 5). The decreased intensity of the N+1 peak in these reactions suggests that transcription of the d-isoguanosine is slowing the enzymes processivity.

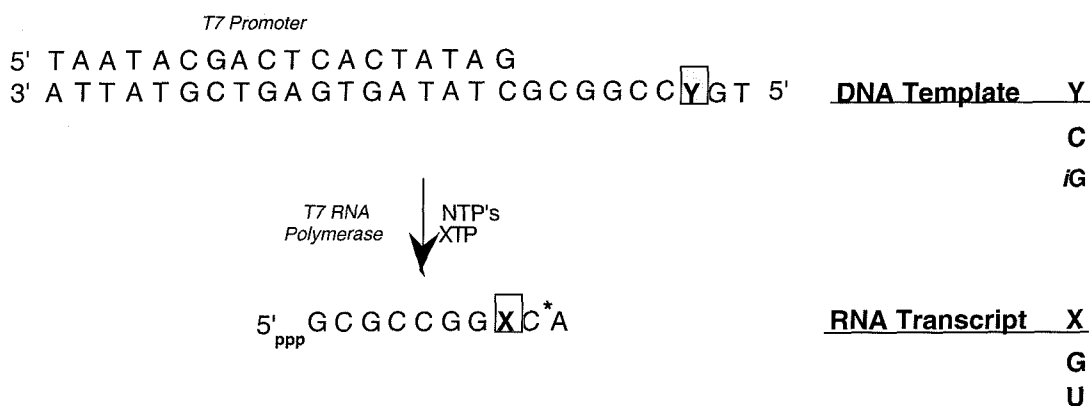


Figure 9. Design of the transcriptional system for the incorporation of $\text{Me}_{\text{iso}}\text{CTP}$ opposite d-isoG.

We were interested in examining the mechanism for the lack of full length products in the presence of MeisoCTP and the absence of UTP. The reactions were body labeled rather than end labeled in order to differentiate between two possible mechanisms for lack of full-length transcription, lack of incorporation or lack of elongation. It is known that

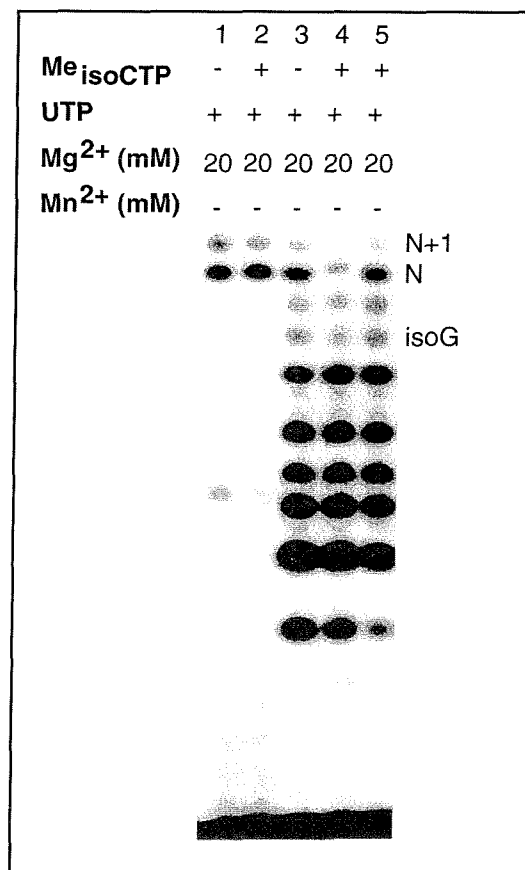


Figure 10. Transcription of a DNA template containing d-isoG, according to Figure 9. The template used for lanes 1 and 2 does not contain d-isoG, Y=C. Lanes 1 and 2 are end labeled with α -³²P-ATP. Lanes 3-8 are labeled with α -³²P-GTP. All triphosphates are present at 1 mM. iG indicates the position corresponding to d-isoguanosine in the transcribed template.

deoxynucleosides act as chain terminators in transcription by T7 RNA polymerase.⁵³ The deoxynucleoside is incorporated into the growing chain, but presumably assumes a conformation that prevents addition of the next base. A similar lack of elongation might account for the lack of full-length product in our system. If this mechanism is involved, a band corresponding to incorporation of a single additional nucleotide should be present, corresponding to the position opposite d-isoG in the template, in the presence of Me_{iso}CTP

and absence of UTP. In our experiment, the gel clearly indicates that the major failure band occurs for the position corresponding to the base preceding the isoG, in the presence or absence of Me_{iso}CTP. This result strongly supports that the *Me_{iso}CTP is not incorporated into the growing chain*. As expected from this result, the addition of Mn²⁺ (lanes 6-8), known to aid transcriptional elongation in the presence of deoxynucleotide triphosphates, did not effect the incorporation of Me_{iso}C.

6.5 Conclusions

An efficient synthesis of deoxyisoguanosine phosphoramidite has been developed. The synthesis of the novel nucleotide Me_{iso}CTP proceeded efficiently. The results of transcription experiments indicate that Me_{iso}CTP does not inhibit T7 RNA polymerase. Me_{iso}CTP is not incorporated opposite d-isoG in transcription reactions using T7 RNA polymerase, but UTP is able to mispair with d-isoG. Our failure to observe the incorporation of Me_{iso}C opposite isoG in transcription reactions supports the suggestion of Benner that d-isoG might favor a different tautomer in the enzyme active site. The tautomerization of isoG will be discussed in detail below. This proposed change in the tautomeric preference for d-isoG causes the inhibition of incorporation of either isoCTP or Me_{iso}CTP, while allowing for "misincorporation" of UTP.

The failure of the transcription of d-isoG containing templates raises interesting evolutionary issues. The lack of incorporation could reflect the inherent instability of the base pairing combination, or reflect the evolutionary preference for the common bases. The potential for the enzyme to enforce a specific tautomeric preference is intriguing. The development of an early genetic code could have drawn on a variety of base pairs with unique hydrogen bonding arrays.³⁴ The preference for the natural base pairs might reflect the tautomeric stability of these bases, consequently limiting mutations due to mispairing

with minor tautomeric forms. Alternatively, the lack of enzymatic recognition of the d-isoG-isoC may reflect the years of evolutionary preference for the natural bases. In either case, the possibility remains that an enzyme can be discovered or engineered which will transcribe the isoC-isoG base pair in both directions. In the absence of such an enzyme, the isoC-isoG system can still be used with these shortcomings for some laboratory applications.

The unique presentation of hydrogen bond donors and acceptors of the isocytosine-isoguanosine base pair raises the issue: why are isoC and isoG absent from the current genetic code.³³ While the isoC-isoG base pair has proved amenable to some laboratory applications,

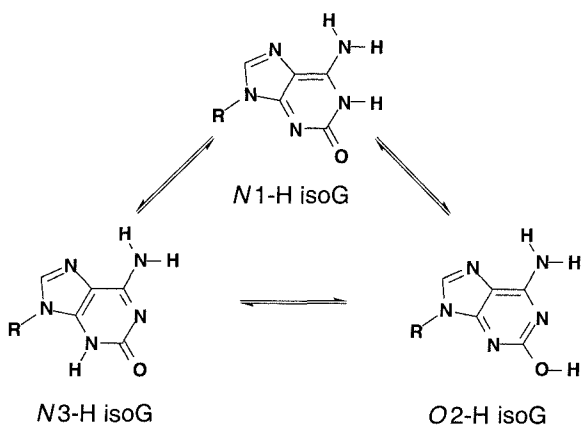


Figure 11. Proposed tautomeric equilibrium of isoguanosine.

the potential tautomerism of isoguanosine has complicated analysis of these experiments. Three reasonable tautomers can describe the structure of isoguanosine (figure 11). Physical organic studies, UV and IR spectroscopy, of the tautomerism of isoguanine shows it to be

predominantly in the *N1-H* form in aqueous solution, with the *O2-H*, enol form favored in less polar solvents.⁵⁴ Low level calculations on isoguanine and isocytidine suggest that tautomeric instability of both isoC and isoG makes them unsuitable for stable storage of genetic information.⁵⁵ Interestingly, in the absence of direct spectroscopic evidence, at least two tautomers of isoguanosine appear to be accessible under biological conditions. Both deoxyisocytosine³⁴ and deoxymethylisocytosine¹ direct the enzymatic incorporation of isoguanosine at the appropriate position in the RNA transcript using T7 RNA polymerase. The fidelity of transcription argues convincingly that the *N1-H* tautomeric form

predominates for isoGTP and *N*⁶-alkyl derivatives of isoGTP.¹ Recently, the enzymatic recognition of the isoC-isoG base pair has been probed in the reverse direction. When transcription is directed by a DNA template containing d-isoG, no incorporation of isoCTP³⁴ or Me_{iso}CTP (this study) is observed. Instead, T7 RNA polymerase prefers to pair d-isoG with UTP. Although the

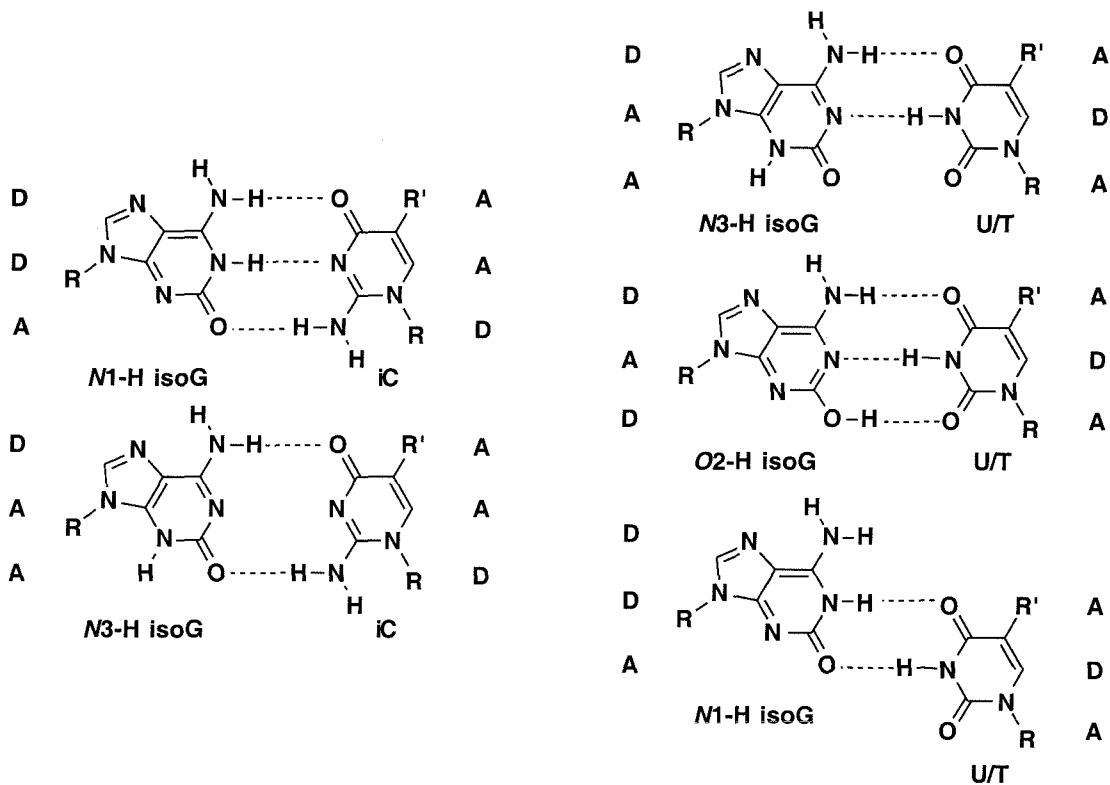


Figure 12. Potential base pairing interactions of the various tautomeric forms of isoguanosine with either isocytosine or thymidine. R=ribo or deoxyribofuranosyl. R'=H for uracil or Me for thymidine. Hydrogen bonding properties are indicated by A and D, acceptor and donor, respectively.

incorporation of UTP may reflect the enzyme's preference for the natural triphosphate, the efficiency of mispairing suggests that the isoG may be present in a different tautomeric form. While Benner argues that the *O*2-H tautomer is probably responsible for the misincorporation,³⁴ both the *O*2-H and *N*3-tautomeric forms of isoG could lead to misincorporation of U (figure 12). Benner's group has also shown that Klenow fragment DNA polymerase, when confronted with a template containing d-isoG, incorporates T

preferentially over d-isoC, though d-isoC is incorporated in the absence of TTP.³⁴ Again, this result suggests that another tautomeric form, either *O2-H* or *N3-H* is accessible.

While tautomerism of the naturally occurring nucleic acid bases has been implicated in mutagenesis,⁵⁶ little evidence is available for the presence of minor tautomeric forms of these bases under biologically relevant conditions.⁵⁷ In structural studies of natural bases, mismatched pairs are always observed to be in *wobble conformations* or in *ionized states* rather than in Watson-Crick geometries with minor tautomeric forms (Morgan, 1993, and references therein). A recent NMR study of the non-natural base pair between *N*⁴-methoxycytosine with guanosine showed two structurally distinct tautomeric forms of the non-natural base,⁵⁹ indicating that higher energy tautomers can be stabilized by favorable base pairing interactions within a DNA duplex. Analogously, the complex *tautomeric equilibrium* of isoguanosine may prove to be dependent on stabilization provided by *intermolecular base-pairing interactions*. Since this work was completed, a crystal structure has been solved that contains two distinct tautomeric forms of d-isoG paired with thymine in duplex DNA, providing direct support for the hypothesis of this study.⁵⁹ We had also attempted to grow crystals of DNA duplexes containing d-isoG paired with d-^{Me}isoC, and had some success, but were not able to obtain crystals that gave soluble diffraction patterns.

6.6 Experimental Section

General. NMR spectra were recorded on a GE 300 instrument operating at 300 MHz. Chemical shifts reported are relative to the solvent residual signal in ppm. UV spectra were recorded on a Hewlett-Packard 8452A diode array spectrometer. Mass spectrometry was performed at the Mass Spectrometry Laboratory at the University of California, Riverside. HPLC analysis was carried out on a HP 1090 analytical HPLC, using a Vydac C18 reversed phase column, (0.46 x 25 cm, 5 μ m TP silica packing).

Preparative HPLC was carried out using a Waters 600e with a Delta-Pak C18 column (Waters). Flash column chromatography was carried out using silica gel 60 (230-400 mesh, Merck). Thin layer chromatography was performed on silica gel 60 F254 plates (Merck). Reagent grade chemicals were used as received in all cases, unless otherwise indicated. Acetonitrile was distilled from calcium hydride. Anhydrous solvents (over 4 Å molecular sieves) were purchased from Fluka.

IsoGTP and Aminoethylisoguanosine triphosphate were synthesized by Dr. Y. Tor, according to the method described.¹ Ribothymidine was synthesized by a known procedure.⁵⁰ The protection of d-isoguanosine was performed according to Seela.⁴⁷

N⁶-(6-Aminopropyl)isoguanosine-5'-monophosphate barium salt (2). 6-methylthioxanthosine-5'-monophosphate triethylammonium salt, prepared according to Tor; Dervan¹ (133 mg, 0.270 mmol) was dissolved in 25 mL of methanol, 25 mL. 33 equiv. of 1,3-diaminopropane (7.45 mL, 8.91 mmol) was added slowly to the solution. The solution was refluxed under Ar for 14 hours, followed by UV spectroscopy. The reaction was concentrated then dissolved in 50 mL of H₂O. Barium acetate, 0.206 g, was added and the mixture stirred for 2 hours, after which the product was precipitated by the addition of 200 mL of ethanol. The ethanol mixture was stirred for 24 hours followed by centrifugation, yielding 141 mg of product (82%). UV(H₂O) λ_{max} 250, 286 nm. ¹H NMR (D₂O) δ 8.16 (s, 1H, H8), 5.94 (d, J = 6.1Hz, 1H, H1'), 4.69 (m, 1H), 4.45 (m, 1H), 4.29 (m, 1H), 3.94 (m, 2H, H5'), 2.78 (m, 4H, CH₂N), 1.70 (m, 2H, CH₂).

N⁶-(6-Trifluoroacetylamidopropyl)isoguanosine-5'-monophosphate-triethylammonium salt (3). Trifluoroacetylation of compound **2** was performed according to a known procedure (Gebeyehu *et al.*, 1987). Purification on a Sephadex A-25 column (1.5x20 cm) was performed by washing column with 300 mL of 50 mM TEAB,

pH 7.5, then eluting the product with a gradient from 50-500 mM TEAB, pH 7.5. Coevaporation with acetonitrile yielded a white product. UV(H₂O) λ_{max} 250, 294 nm. ¹H NMR (CD₃OD) δ 7.98 (s, 1H, H8), 5.74 (d, J=6.3Hz, 1H, H1'), 4.41 (t, J = 5.7 Hz, 1H), 4.25 (m, 1H), 4.12 (br m, 1H), 4.00 (br d, 2H, H5'), 3.35-2.85 (m, CH₂N), 1.88-1.79 (m, 2H, CH₂), 1.21-1.12 (m, CH₃, Et₃N).

N⁶-(6-Trifluoroacetylamidopropyl)isoguanosine-5'-triphosphate

triethylammonium salt (4). N⁶-(6-Trifluoroacetylamidopropyl)isoguanosine-5'-monophosphate-triethylammonium salt (**3**) 50 mg (0.08 mmol) was dissolved in methanol and repeatedly coevaporated with acetonitrile, then dried under high vacuum for several hours. The starting material was dissolved in 1.75 mL of dry DMF. To this solution was added a solution of 62.5 mg (0.375 mmol) of 1,1'-carbonyldiimidazole in DMF (0.75 mL). The solution gradually became clear. After 2.5 hours, 8 eq. of MeOH were added (25 μ L). The solution was stirred for 30 minutes followed by the addition of 156 mg of tributylammonium pyrophosphate (Sigma) in 1.0 mL dry DMF. After 1 hour another 100 μ L of the tributylammonium solution was added. After stirring for 48 hours at room temperature, the reaction mixture was filtered and concentrated. The residue was dissolved in 100 mM TEAB pH 7.5 (25 mL) and loaded on a Sephadex A-25 column equilibrated with the same buffer at 4 °C. The column was eluted with a gradient from 0.1 M to 1.0 M TEAB pH 7.5. Pure fractions were combined and evaporated to yield 26 mg (63%). UV(H₂O) λ_{max} 250, 294 nm. ¹H NMR (D₂O) δ 8.18 (s, 1H, H8), 5.91 (d, J = 6.0 Hz, 1H, H1'), 4.68 (br m, 1H), 4.54 (br m, 1H), 4.34 (br m, 2H, H5'), 3.45 (br m, 2H, CH₂N), 3.31-2.98 (m, CH₂N, including Et₃N), 1.63 (br m, 2H, CH₂), 1.37-1.17 (m, CH₃, Et₃N).

N⁶-(6-Aminopropyl)isoguanosine-5'-triphosphate-triethylammonium salt

(5). The protected amine (**4**) 30 mg, was treated with 1N NaOH to give a pH of approximately 11. The solution was stirred for 12 hours. The pH was adjusted with TEAB to 7.5. Purification on a Sephadex A-25 column (1.5x20 cm) was performed by washing column with 300 mL of 50 mM TEAB, pH 7.5 then eluting the product with a gradient from 50-500 mM TEAB, pH 7.5. Coevaporation with acetonitrile yielded a white product, 26 mg (75%). UV(H₂O) λ_{max} 250, 294 nm. ¹H NMR (D₂O) δ 8.24 (br s, 1H, H8), 5.91 (d, 1H, H1'), 4.68 (br m, 1H, H2'), 4.53 (br m, 1H, H3'), 4.35 (br m, 1H, H4'), 4.21 (br m, 2H, H5'), 3.52-2.97 (m, CH₂, including Et₃N), 2.04-1.67 (br m, CH₂), 1.37-1.22 (m, CH₃ [Et₃N]).

2'-Deoxy-adenosine N¹-oxide (7). 4.0 g (14.88 mmol) of 2'-deoxyadenosine (Sigma) was dissolved in water (110 mL). An ether solution of monopermaleic acid (47.6 mmol, 3.2 eq.), generated according to Mantsch *et al.*; 1975, was added to 50 mL of water. The ether was removed by stirring under an air stream. The aqueous solution of the monopermaleic acid was chilled in an ice bath and the pH adjusted to approximately 7 with 1N NaOH. The monopermaleic acid solution was added to the deoxyadenosine solution and stirred for 36 hours at room temperature under aluminum foil. The reaction was then acidified to pH 4.5 with 1N HCl and added to 800 mL of EtOH. Maleic acid was precipitated from the reaction mixture. The reaction was filtered and concentrated. Flash column chromatography (10-20% H₂O in CH₃CN) yielded pure white product. 3.6g (85%). ¹H NMR (D₂O) 8.56 (s, 1H), 8.40 (s, 1H), 6.46 (t, J = 6.6 Hz, 1H, H1'), 4.63 (m, 1H), 4.13 (m, 1H), 3.77 (m, 2H), 2.85 (m, 1H), 2.60-2.52 (m, 1H).

2'-Deoxyisoguanosine (8). 2.56 g of 2'-deoxyadenosine-N¹-oxide **7** was dissolved in 900 mL of 10 mM NH₄OH. The resulting solution was irradiated for 48 hours using a

Hg lamp equipped with a mechanical stirring device. After irradiation, the reaction was concentrated *in vacuo*. The concentrated mixture (approximately 100 mL) was slurried with silica gel, and repeatedly coevaporated with acetonitrile until the silica was dry. The silica was then slurried with acetonitrile and loaded onto a prepoured flash column equilibrated with 10% H₂O/CH₃CN. The column was washed with 10% H₂O/CH₃CN until no dA was eluted then the d-isoguanosine was eluted with 20% H₂O/CH₃CN, yielding 1.12g. (45%). ¹H NMR was in agreement with that previously reported.⁴⁷

6-((Dimethylamino)methylidene)isoguanosine (9). 2'-deoxyisoguanosine was protected according to the literature method to give the desired product, **9**. ¹H NMR was in agreement with that previously reported.⁴⁷

5'-4,4'-Dimethoxytrityl-6-((dimethylamino)methylidene)isoguanosine (10). Compound **9** was reacted with DMT-Cl according to the literature to yield the desired product. ¹H NMR was in agreement with that reported.⁴⁷

5'-O-4,4'-Dimethoxytrityl 6((dimethylamino)methylidene)isoguanosine 3' - [(2-cyanoethyl) N,N-diisopropylphosphoramidite] (11). Phosphitylation was carried out with the tetrazole/ diphosphoramidite method. 0.5 g of **10** was dried by repeated coevaporation from dichloromethane with acetonitrile. 50 mg of tetrazole (Aldrich, sublimed). Dissolved in 5 mL of dry CH₂Cl₂ and 7.5 mL dry CH₃CN. Cyanoethyl N,N,N',N'-tetraisopropylphosphoramidite (260 μL, 1.05 eq.) was added slowly to the solution and stirred for 2 hours at room temperature. After 2 hours, 0.5 equivalents of phosphitylating agent (125 μL) were added slowly and the reaction stirred for another hour. The reaction was concentrated, dissolved in CH₂Cl₂ with 0.5% Et₃N and purified by flash chromatography (0-5% MeOH/ 0.5% Et₃N/CH₂Cl₂) yielding 502

mg (60%). ^1H -NMR (CDCl_3) δ 8.58 (s, 1H, H8), 7.25 (s, 1H), 7.12 (d, $J=6.8\text{Hz}$, 1H), 6.72-6.40 (br m, 12 H, DMT), 6.15-6.02 (br m, 4H, DMT), 5.64 (m, 1H), 5.16(s, 1H, H1'), 3.85 (br m, 1H), 3.52-3.33 (br m, 6H), 3.17 (m, 1H), 2.90-2.73 (m, 6H), 2.48, 2.41 (s, 3H, NCH_3), 2.38-2.17 (m, 3H), 2.09-1.95 (m, 6H), 0.65-0.32 (m, 12H).

2, 5'- Anhydro- 5-methyluridine (13). According to Shibuya,⁵¹ 500 mg of ribothymidine, **12**, was dissolved in 3.8 mL of anhydrous THF with 1.52 g of triphenylphosphine. Diethylazodicarboxylate (DEAD) was added slowly as a solution of 0.93 mL (5.8 mmol) in 1 mL of THF. The reaction was stirred overnight, during which time a white precipitate formed. The precipitate was collected by centrifugation and washed with ether to yield 720 mg (76%). ^1H NMR ($\text{DMSO}-d_6$) δ 7.94 (s, 1H, H4), 7.46-7.26 (m, 15H, C_6H_5), 5.76 (s, 1H, H1'), 4.77 (q, 1H, H4'), 4.69 (s, 1H, H3'), 4.50-4.46 (m, 2H, H5'), 4.08 (d, 1H, H2'), 1.76 (s, 3H, Me).

5-Methylisocytosine (14). 0.5 g of 2,5'-anhydro-5-methyluridine **13** was suspended in 25 mL of anhydrous MeOH in a sealed tube under Argon. The tube was saturated with $\text{NH}_3(\text{g})$ in an ice bath. The tube was then heated at 100°C for 4 hours with stirring. The reaction was then concentrated to approximately 1 mL and precipitated with 50 mL of acetonitrile yielding 195 mg, 76%. ^1H NMR (CD_3OD) δ 5.42 (d, $J = 6.7\text{ Hz}$, 1H, H1'), 4.36 (t, $J = 6.1\text{ Hz}$, 1H, H2'), 4.14 (m, 1H, H3'), 4.05 (m, 1H, H4'), 3.76 (m, 2H, H5'), 1.84 (s, 3H, Me).

5-Methylisocytosine-5'-triphosphate triethylammonium salt (15). 5-methylisocytosine (**14**) was phosphorylated according to a literature procedure.⁵² $\text{Me}_{\text{iso}}\text{C}$ (100 mg, 0.39 mmol) was suspended in 3.9 mL of trimethyl phosphate and chilled in an ice bath to 0°C . The solution was treated with phosphorine chloride (117 μL , 1.17 mmol).

The reaction was stirred at 0 °C for 3 hours followed by the simultaneous addition of 7.8 mL of a 0.5 M solution of tributylammonium pyrophosphate in DMF and 1.17 mL of tributylamine. After 3 minutes, the reaction mixture was transferred to a flask containing 30 mL of 1.0 M TEAB pH 7.7. The mixture was evaporated and coevaporated with MeOH then with acetonitrile. The product was purified by chromatography using Sephadex A-25. The crude mixture was loaded on the column in 100 mM TEAB, pH 7.5, and eluted with a gradient of 0.1 - 1.0 M TEAB pH 7.5 over 1.5 L, yielding 37 mg (17%). UV (H₂O) λ_{max} 260 nm. ¹H NMR (D₂O) δ 7.60 (s, 1H, H4), 5.45 (d, J = 6.4 Hz, 1H, H1'), 4.49 (m, 1H), 4.40 (m, 1H), 4.27-4.08 (m, 3H), 3.59 (d, J = 7.4 Hz, 1H), 3.28-2.91 (m, CH₂N), 1.84 (s, 3H, Me), 1.42-0.80 (m, CH₃, Et₃N).

Synthesis and purification of oligonucleotides. Automated DNA and RNA synthesis was performed on an Applied Biosystems 394b DNA synthesizer using 2-cyanoethyl phosphoramidites (Applied Biosystems or Glen Research). All other reagents used for oligonucleotide synthesis were obtained from either ABI or Glen Research. All novel phosphoramidites were used as 0.1M solutions in acetonitrile (Fluka, over 4 Å sieves). RNA oligonucleotides were deprotected according to the procedure recommended by ABI. Oligonucleotides were purified on 15-20% preparative polyacrylamide gels and isolated by UV-shadowing. The isolated bands were extracted through a crush and soak procedure in 1X TBE buffer, or by electroelution (Schleicher; Schuell). The oligonucleotides were desalted either by Sep-pak cartridges (Waters) or by extensive dialysis against Milli-Q water (Waters). Concentrations of oligonucleotides were determined using the following molar extinction coefficients for each base at 260 nm: 15400 (A), 11700 (G), 7300 (C), 8800 (T), 10100 (U), 6300 (MeisoC), 4600 (isoG and N6 substituted isoG derivatives) and 5000 (d-isoG). Digestion and HPLC analysis was

performed for the deoxyoligonucleotides containing non-natural nucleosides as previously described.¹

Transcription reactions. Small scale transcription experiments were performed according to known procedures.⁵³ The single stranded templates were annealed to the 18-mer promoter in 100 mM NaCl-TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.8) by heating a 1:1 mixture to 90°C for 2 minutes followed by cooling to room temperature over a period of approximately 1 hour. Solutions of the annealed promoter were made to give a final concentration of 5 μ M and stored at -20 °C. Small scale transcription reactions were performed in 20 μ L volumes containing 250 nM template, 120 units T7 RNA polymerase (Pharmacia), 40 mM Tris-HCl pH 8.1 (at 37 °C), 20 mM MgCl₂, 5 mM DTT, 1 mM spermidine hydrochloride, 0.01% Triton X-100, 50 μ g/mL BSA (Pharmacia, RNase/DNase free), 4 mg/mL PEG 8000, RNase Inhibitor (Boehringer Mannheim), 1 mM NTP's and 1-10 μ Ci labeled triphosphate (NEN). The reactions were quenched after 3 hours by the addition of 50 μ L of loading buffer (7 M urea, 1 mM EDTA, pH 7.8). The reactions were heated briefly at 75 °C then loaded on 20% polyacrylamide gels for analysis. Large scale transcription experiments were performed as above with the addition of 5 mM DTT. Large scale reactions were centrifuged to remove magnesium pyrophosphate then ethanol precipitated using glycogen (300 μ g) as a carrier. The RNA was resuspended in loading buffer (7 M urea, 1x TE, pH 7.8) and purified on 15% polyacrylamide gels. The product bands were excised after UV shadowing and crushed. Soaking overnight in 0.5 M ammonium acetate, pH 5.2, overnight was followed by filtration and ethanol precipitation of the RNA. 5'-Labeling of the transcription products and of DNA oligonucleotides was performed by 5'-phosphorylation using T4 polynucleotide kinase (Boehringer Mannheim) and γ -³²P-ATP.

Hammerhead kinetics. Kinetic experiments with the hammerhead ribozyme were carried out according to the literature method.³⁷ ³²P-5'-Labeled substrate was

annealed to the ribozyme strand by briefly heating to 75 °C (1 minute), followed by cooling to 37 °C. The reactions were initiated by the addition of Mg^{2+} to a final concentration of 10 mM. Aliquots were taken from the reaction and quenched into loading buffer containing 50 mM EDTA followed by rapid vortexing and chilling to 0 °C. The cleaved fragments were then separated on a 20% polyacrylamide gel and quantified by phosphorimaging. Four independent runs were averaged for each mutant studied.

Fe-EDTA Autocleavage. Attempts to autocleave RNA containing AH-iG-EDTA were carried out by preincubation of the RNA with 10 mM Mg^{2+} with heating to 75 °C followed by slow cooling to 37 °C. The RNA was then incubated with Fe^{2+} for 2 hours at 37 °C. Cleavage reactions were initiated with DTT (4 mM) followed by incubation at 37 °C for 6 to 12 hours. Reactions were quenched with an equal volume of 1 X TBE, pH 7.8, 50 mM EDTA. Products were separated on a 20% polyacrylamide gel.

Purification of Oligonucleotides for Structural Studies. Oligonucleotides for crystallization were synthesized by the procedures outlined above, with the exception that the 5' trityl group was not removed. Trityl on purification was performed by HPLC (Waters 600E) using a reversed phase column (Delta-Pak C18, Waters) with a gradient from 15-45% CH_3CN in 100 mM TEAA, pH 6. The oligonucleotides were lyophilized and treated with 50% acetic acid for 20 minutes to remove the trityl groups. Following detritylation, the oligonucleotides were lyophilized and repurified on the same column using a shallow gradient (0-20% CH_3CN in 100 mM TEAA, pH 6 over 45 minutes. The oligonucleotides were lyophilized then desalted (Sep-pak, Waters). The desalted oligonucleotides were subjected to cation exchange chromatography (Sephadex, CM25 (Na^+ form)) to replace the triethylammonium counterions.

The following sequences were prepared:

Table 2. Sequences for crystallization.

Template	5'	C	C	A	A	C	G	T	T	G	G	3'
1		C	C	A	A	<u>iC</u>	<u>iG</u>	T	T	G	G	
2		C	C	A		<u>iG</u>	T	A	<u>iC</u>	T	G	G
3		C	C	A	A	<u>I</u>	<u>iG</u>	T	T	G	G	
4		C	C	A		<u>iG</u>	T	A	<u>I</u>	T	G	G

Successful crystallizations follow the discovery of "template" sequences.⁶⁰ In collaboration with the Rees group, specifically with Dr. Joshua-Tor, we are attempting to crystallize DNA duplexes containing the non-natural base pairs of isoguanosine. With this in mind, a template was chosen which has been crystallized with a wide variation of internal sequences. The sequences we are attempting to crystallize, shown in table 2, are derived from the sequence crystallized by Dickerson and coworkers, 5' CCAACGTTGG-3'.⁶¹ This template has been crystallized with a variety of internal sequences, including a GA mismatch⁶² and a C-I base pair.⁶³ In order to probe the structure of this novel base pair, we are attempted to crystallize four self complementary decamer sequences, two with isoC-isoG base pairs, and two with isoG-T base pairs, table 2. Sequence 4 gave highly diffracting crystals, but with insoluble diffraction patterns.

6.7 References

- (1) Tor, Y.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 4461-4467.
- (2) Guerrier-Takada, C; Gardiner, K; Marsh, T; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849-857.
- (3) Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. *Cell* **1982**, *31*, 147-154.
- (4) Jackson, R.J. *Cell* **1993**, *74*, 9-14.
- (5) Cech T.R. *Ann. Rev. Biochem.* **1990**, *59*, 543 -568.
- (6) Symons, R.H. *Ann. Rev. Biochem.* **1992**, *61*, 641 -671.
- (7) Abelson, J. *Harvey Lect.* **1991**, *85*, 1 -42.
- (8) Noller, H. F.; Hoffarth, V.; Zimniak, L. *Science* **1992**, *256*, 1416-1419.
- (9) Beaudry, A.A.; Joyce, G.F. *Science* **1992**, *257*, 635-641.
- (10) Vaishnav, Y.N.; Wongstaal, F. *Annu. Rev. Bioch.* **1991**, *60*, 577 -630.
- (11) Lima, W.F.; Monia, B.P.; Ecker, D.J.; Freier, S.M. *Biochemistry* **1992**, *31*, 12055-12061.
- (12) Quigley and Rich, 1976.
- (13) Holbrook, S.R.; Cheong, C.J.; Tinoco, I. *Nature* **1991**, *353*, 579-581.
- (14) Brown, R.S.; Dewan, J.C.; Klug, A. *Biochemistry* **1985**, *24*, 4785-4801.
- (15) Scott, W. G.; Finch, J. T.; Klug, A. *Cell* **1995**, *81*, 991-1002.
- (16) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K. H.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678-1685.
- (17) Varani, G; Tinoco, I. *Q. Rev-Bioph.* **1991**, *24*, 479 -532.
- (18) Nikonowicz, E.P.; Pardi, A. *Nature* **1992**, *355*, 184 -186.
- (19) Doudna, J.A.; Grosshans, C.; Gooding, A.; Kundrot, C.E. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7829 -7833.
- (20) Draper, D.E. *Acc. Chem. Res.* **1993**, *25*, 201-207.
- (21) Ehresman, C.; Baudin, F.; Mougél, M.; Romby, P.; Ebel, J.; Ehresman, B. *Nucleic Acids Res.* **1987**, *15*, 53-71.

- (22) Han, K.; Kim, H.J. *Nucleic Acids Res.* **1993**, *21*, 1251-1257; Zuker, M.; Jaeger, J.A.; Turner, D.H. *Nucleic Acid Res.* **1991**, *19*, 2707-2714.
- (23) Gautheret, D.; Cedergren, R. *FASEB J.* **1993**, *7*, 97 -105.
- (24) Michel, F.; Westhof, E. *J. Mol. Biol.* **1990**, *216*, 585-610.
- (25) Hertzberg, R.P.; Dervan, P.B. *J. Am. Chem. Soc.* **1982**, *104*, 313-315; Vary, C.P.H.; Vournakis, J.H. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 6978-6982; Kean, J.M.; White, S.A.; Draper, D.E. *Biochemistry* **1985**, *24*, 5062-5070.
- (26) Celander, D.W.; Cech, T.R. *Science* **1992**, *252*, 401 -407; Latham J.A.; Cech T.R. *Science* **1989**, *245*, 276 -282; Murphy, F.L.; Cech, T.R. *Biochem* **1993**, *32*, 5291 -5300.
- (27) Chow, C.S.; Behlen, L.S.; Uhlenbeck, O.C.; Barton, J.K. *Biochemistry* **1992**, *31*, 972-982.
- (28) Wang, J.F.; Cech, T.R. *Science* **1992**, *256*, 526 -529.
- (29) Han, H. Y.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4955-4959.
- (30) Beaucage, S.L.; Iyer, R.P. *Tetrahedron* **1993**, *49*, 1925-1963.
- (31) Reyes, V.M.; Abelson, J.N. *Methods Enzym.* **1989**, *180*, 63-69; Yisraeli, J.K.; Melton, D.A. *Methods Enzym.* **1989**, *180*, 42-50.
- (32) Moore, M.J.; Sharp, P.A. *Science* **1992**, *256*, 992-997.
- (33) Rich, A.; In *Horizons in Biochemistry*, Kasha, M.; Pullman, B.; Eds; Academic Press: New York, 1962, 103-126.
- (34) Switzer, C.; Moroney, S.E.; Benner, S.A. *J. Am. Chem. Soc.* **1989**, *111*, 8322-8323; Switzer, C.Y.; Moroney, S.E.; Benner, S.A. *Biochemistry* **1993**, *32*, 10489-10496.
- (35) Bain, J.D.; Switzer, C.; Chamberlin, A.R.; Benner, S.A. *Nature* **1992**, *356*, 537-539.
- (36) Uhlenbeck, O.C. *Nature* **1987**, *328*, 596-600.
- (37) Ruffner, D.E.; Stormo, G.D.; Uhlenbeck, O.C. *Biochemistry* **1990**, *29*, 10695-10702.
- (38) Dahm, S.; Uhlenbeck, O.C. *Biochemistry* **1991**, *30*, 9464-9469; Dahm, S.C.; Derrick, W.B.; Uhlenbeck, O.C. *Biochemistry* **1993**, *32*, 13040-13045.
- (39) Heus, H.A.; Uhlenbeck, O.C.; Pardi, A. *Nucleic Acids Res.* **1990**, *18*, 1103-1108.
- (40) Fedor, M.J.; Uhlenbeck, O.C. *Biochemistry* **1992**, *31*, 12042-12054.

- (41) Fu, D.; McLaughlin, L.W. *Biochemistry* **1992**, *31*, 10941-10949; Fu, D.; McLaughlin, L.W. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3985-3989.
- (42) Tuschl, T.; Ng, M.; Pieken, W.; Benseler, F.; Eckstein, F. *Biochem* **1993**, *32*, 11658-11668.
- (43) Woisard, A.; Favre, A.; Clivio, P.; Fourrey, J. *J. Am. Chem. Soc.* **1992**, *114*, 10072-10074.
- (44) Dreyer, G.B.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 968-972.
- (45) Han, H. Y.; Dervan, P. B. *Nucl. Acids Res.* **1994**, *22*, 2837-2844.
- (46) Tor, Y.; Dervan, P.B.; unpublished results.
- (47) Seela, F.; Mertens, R.; Kazimierczuk, Z. *Helv. Chim. Acta* **1992**, *75*, 2298-2306.
- (48) Cramer, F.; Schlingloff, G. *Tetrahedron Lett.* **1964**, 3201-.
- (49) Roberts, C.; Bandaru, R.; Switzer, C. *J. Am. Chem. Soc.* **1997**, *119*, 4640-4649; Roberts, C.; Bandaru, R.; Switzer, C. *Tet. Lett.* **1995**, *36*, 3601-3604.
- (50) Mansuri, M.M.; Starrett, J.E.; Wos, J.E.; Tortolani, D.R.; Brodfuehrer, P.R.; Howell, H.G.; Martin, J.C. *J. Org. Chem.* **1989**, *54*, 4780-4785.
- (51) Shibuya, S.; Kuninaka, A.; Yoshino, H. *Chem. Pharm. Bull.* **1974**, *22*, 719-721.
- (52) Mishra, N.C.; Broom, A.D. *J. Chem. Soc. Chem. Comm.* **1991**, 1276-77.
- (53) Milligan, J.F.; Groebe, D.R.; Witherell, G.W.; Uhlenbeck, O.C. *Nucleic Acids Res.* **1987**, *15*, 8793-8798; Milligan, J.F.; Uhlenbeck, O.C. *Methods Enzym.* **1989**, *180*, 51-62.
- (54) Sepiol, J; Kazimierczuk, Z.; Shugar, D. *Z. Naturforsch* **1976**, *31C*, 361-370; Shugar, D.; Kierdaszuk, B. *J. Biosci.* **1985**, *8*, 657-668.
- (55) Jaworski, A.; Kwiatkowski, J.S.; Lesyng, B. *Int. J. Quant. Chem.: Quant. Biol. Symp.* **1986**, *12*, 209-216.
- (56) Watson J.; Crick, F. *Nature* **1953**, *171*, 964-967.
- (57) Morgan, A.R. *Trends Bioch. Sci.* **1993**, *18*, 1-3.
- (58) Fazakerley, G.V.; Gdabiec, Z.; Sowers, L.C. *J. Mol. Biol.* **1993**, *230*, 6-10.
- (59) Robinson, H.; Gao, Y. G.; Bauer, C.; Roberts, C.; Switzer, C.; Wang, A. H. J. *Biochem.* **1998**, *37*, 10897-10905.
- (60) Joshua-Tor, L.; Sussman, J.L. (1993) *Curr. Opin. Structr. Biol.* **1993**, *3*, 323-335.
- (61) Grzeskowiak, K; Yanagi, K; Prove, G.G.; Dickerson, R.E. *J. Biol. Chem.* **1991**, *266*, 8861- 8883.
- (62) Prive, G.G; Yanage, K.; Dickerson, R.E. *J. Mol. Biol.* **1991**, *217*, 177-199.

- (63) Lipanov, A.; Kopka, M.L; Kaczorgzeskowiak, M; Quintana, J.; Dickerson, R.E.
Biochemistry **1993**, 32, 1373-1389.