

SYNTHESIS AND STRUCTURAL STUDIES OF CYCLIC Py-Im POLYAMIDES

Thesis by  
David Michael Chenoweth

In Partial Fulfillment of the Requirements  
for the Degree of  
Doctor of Philosophy

California Institute of Technology

Pasadena, California

2009

(Defended January 15, 2009)

© 2009

David Michael Chenoweth

All Rights Reserved

*...for Kimberly, my love...*

## Acknowledgments

This thesis is dedicated to my wife Kim for her constant love, support, and friendship through the years. I can't thank her enough for her love and support, which has guided me through our time at Caltech. Our scientific collaborations and discussions have been another source of constant enjoyment for me. This thesis would not have been possible without her help and support and I am filled with excitement as I think about the arrival of our first child and our future together.

I would like to thank my Ph.D. committee Peter Dervan, Linda Hsieh-Wilson, Doug Rees, and Brian Stoltz for all the time and effort they have invested in me while at Caltech. They have all been fantastic role models and their dedication and passion for science has been truly inspiring. I would like to express my sincere gratitude to my advisor Peter B. Dervan for his guidance, support, and constant willingness to let me pursue all my crazy ideas and projects while in his group. I have truly valued his guidance and mentoring as I navigated my way through research and I thank him for the privilege to work in his research group. I sincerely thank the chair of my committee Linda Hsieh-Wilson for all of her guidance and support during graduate school. Teaching Chem 145 with her has been one of the highlights of my time at Caltech. I will miss the enjoyment of trying to make the best possible problem sets and test questions together. I think often times I learned more than the students during our endeavors. Doug Rees has been a fantastic mentor and I am grateful for his support throughout my Ph.D, his excitement about science, and his mentorship has profoundly influenced me and will continue to be an inspiration. My interactions with Doug and his research group have been one of the highlights of my graduate school experience at Caltech. I thank Brian Stoltz for serving on my committee. His passion for science, intellectual curiosity, and hard work are truly inspiring. The time he has devoted and his comments, honest criticisms, and mentoring during my proposal exams and thesis defense are greatly appreciated.

Many other professors have made an incredible impression on me while at Caltech and I thank them all. Bill Goddard has been a fantastic collaborator and mentor and his excitement about science is truly infectious. I will miss our discussions about science and collaborations on projects orthogonal to my Ph.D research. His generosity and willingness to give me access to his computers and group resources has been very valuable and is greatly appreciated. I thank Harry Gray for his mentorship and advice and for being another inspirational professor to me. I especially valued our discussions about research (fluorescent polyamides) and his willingness to share his group resources. Dennis Dougherty is thanked for his advice and research discussions (lone pair- $\pi$

interactions and electrostatic potential map analysis). Dennis has also been an inspirational professor and his Physical Organic Chemistry textbook has served as a constant source of enjoyment to me, often occupying my night stand and keeping me reading late into the night. Jack Roberts is thanked for our many discussions about NMR and his constant willingness to talk to students and help them with problems. Just walking down the hall between Church and Crellin every day and seeing Jack intensely pounding away on his computer has been very motivating. I always knew he was thinking deeply about science and this was very comforting for me.

I have had the privilege to work with many collaborators while at Caltech and I thank them all. Dan Harki has been a great friend and collaborator and I greatly enjoyed our time at Caltech and look forward to future collaborations. I would also like to thank Christian Dose for many great times in the lab and the rest of my collaborators that I have had the privilege of working with (Justin Cohen, John Phillips, Anne Viger, Mike Marques, Mike Brochu, and Julie P.). I also thank the rest of the Dervan group for making the lab a wonderful place.

I sincerely thank Mike Day and Jens Kaiser for putting up with my constant questions and nagging day in and day out during my adventures in crystallography. I thank both of them for their guidance and friendship during my time at Caltech. I will definitely remember the Friday night plasma party with Jens in the Rees group microwave and the many fun synchrotron trips. I thank Bill Scott at UCSB for help with crystallography software setup on my mac and the Chimera team at UCSF for their help with new software features and script writing.

I sincerely thank all of my friends and family outside Caltech for their constant love and support through the years: Mom, Dad, Dannielle, Ed and Sandy Shuler, and Kieth Shuler, Angela Harki, Sam Harki, and Tom Britton. We really miss all of you! Angela and Sam Harki are thanked for being great friends and listening to all the complaints during Grad school. We will definitely miss the two of you (and Dan too)! We will have to find some way to visit each year.

I would especially like to thank Tom Britton for his friendship and mentoring during my time at Eli Lilly. He has made a lasting impression on me and taught me more than I could imagine.

## Abstract

The work presented in this thesis is focused on the molecular recognition of DNA by minor groove binding polyamides. Methods and strategies for the solution-phase synthesis of hairpin and cyclic pyrrole-imidazole polyamides are presented with optimized protocols requiring little to no chromatography. These synthetic strategies have led to the design of cyclic polyamides targeted to the androgen response element and are shown to be biologically active and cell permeable in cell culture experiments in addition their binding affinities rival that of most polyamide architectures. The structural elucidation of an  $\alpha$ -amino-turn-linked cyclic polyamide is presented at 1.17 Å resolution providing insight into the detailed molecular recognition process and allosteric modulation responsible for the inhibition of transcription factor-DNA binding. Additionally, structural elucidation of a  $\beta$ -amino-turn-linked cyclic polyamide, highlighting the conformational differences compared to the  $\alpha$ -amino-turn linked structure is presented. A structural basis for the inability of polyamides to bind dsRNA is also proposed based on biophysical, structural, and modeling data. In addition to these studies a new class of programmable oligomers targeting the DNA sequence 5'-WGGGGW-3' were shown to inhibit DNA binding of the Nf-kB transcription factor by EMSA gel shift. Compounds synthesized in this study were found to possess unique fluorescent properties with the ability to modulate their fluorescence by binding their targeted dsDNA, leading to sequence specific fluorescent detection reagents. Efforts toward the templated-assembly of polyamides using higher-order DNA structure (NCP) are also reported and the development of a new profluorescent class of heterocycle, which has the potential to be used as a chemical reporter of ligation events is described.

## Table of Contents

Dedication.....	iii
Acknowledgments.....	iv
Abstract.....	vi
Table of Contents.....	vii
List of Figures.....	xii
List of Schemes.....	xx
List of Tables.....	xxi
Nomenclature and Symbology.....	xxiii
Chapter 1: Introduction to Molecular Recognition of DNA.....	1
1.1 Background and Significance.....	2
1.2 Nucleic Acid Structure.....	3
1.3 Molecular Recognition of DNA.....	5
1.5 Scope of this work.....	23
1.6 Notes and Reference.....	25
Chapter 2: Solution-Phase Synthesis of Pyrrole–Imidazole Polyamides.....	30
Abstract.....	31
2.1 Introduction.....	32
2.2 Results and Discussion.....	33
2.3 Conclusion.....	37
2.4 Experimental Section.....	37
2.4.1 General.....	37
2.4.2 HCl•H <sub>2</sub> N-Py-CO <sub>2</sub> Me ( <b>10</b> ).....	39
2.4.3 BocHN-PyPy-CO <sub>2</sub> Me ( <b>11</b> ).....	39
2.4.4 HCl•H <sub>2</sub> N-PyPy-CO <sub>2</sub> Me ( <b>12</b> ).....	39
2.4.5 BocHN-PyPyPy-CO <sub>2</sub> Me ( <b>13</b> ).....	39
2.4.6 HCl•H <sub>2</sub> N-PyPyPy-CO <sub>2</sub> Me ( <b>4</b> ).....	40
2.4.7 ImPyPyPy-CO <sub>2</sub> Me ( <b>14</b> ).....	40
2.4.8 ImPyPyPy-CO <sub>2</sub> H ( <b>3</b> ).....	41
2.4.9 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -Im-CO <sub>2</sub> Et ( <b>15</b> ).....	41
2.4.10 HCl•H <sub>2</sub> N-(R) <sup>β</sup> -CbzHN $\gamma$ -Im-CO <sub>2</sub> Et ( <b>5</b> ).....	42
2.4.11 ImPyPyPy-(R) <sup>β</sup> -CbzHN $\gamma$ -Im-CO <sub>2</sub> Et ( <b>16</b> ).....	42
2.4.12 ImPyPyPy-(R) <sup>β</sup> -CbzHN $\gamma$ -Im-CO <sub>2</sub> H ( <b>17</b> ).....	43
2.4.13 ImPyPyPy-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>2</b> ).....	43
2.4.14 ImPyPyPy-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> H ( <b>22</b> ).....	44
2.4.15 BocHN-(+)- <sup>Bn</sup> OIPA ( <b>20</b> ).....	44
2.4.16 ImPyPyPy-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-(+)- <sup>Bn</sup> OIPA ( <b>23</b> ).....	45
2.4.17 ImPyPyPy-(R) <sup>β</sup> -H <sub>2</sub> N $\gamma$ -ImPyPyPy-(+)-IPA ( <b>1</b> ).....	46
2.4.18 Calculation of Molar Extinction Coefficients.....	47
2.5 Notes and References.....	48
2.6 Spectra and Supplemental Information.....	50
Chapter 3: Cyclic Pyrrole–Imidazole Polyamides Targeted to the Androgen Response Element.....	82
Abstract.....	83
3.1 Introduction.....	84
3.2 Results and Discussion.....	85
3.2.1 Solution-Phase Synthesis of Cyclic Polyamides.....	85
3.2.2 Thermal Stabilization of DNA duplexes by Polyamides.....	87
3.2.3 Biological Assay for Cell Permeability.....	89
3.2.4 ADMET Studies of Polyamides <b>1</b> and <b>5</b> .....	<b>89</b>

3.3 Conclusion .....	90
3.4 Experimental Section .....	90
3.4.1 General .....	90
3.4.2 UV Absorption Spectrophotometry .....	91
3.4.3 Measurement of Androgen-Induced PSA mRNA .....	91
3.4.4 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -Im-CO <sub>2</sub> H ( <b>8</b> ) .....	91
3.4.5 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>9</b> ) .....	92
3.4.6 HCl•H <sub>2</sub> N-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>10</b> ) .....	92
3.4.7 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> H ( <b>11</b> ) .....	93
3.4.8 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>12</b> ) .....	93
3.4.9 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> H ( <b>13</b> ) .....	94
3.4.10 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> Pfp ( <b>14</b> ) .....	95
3.4.11 cyclo-(-ImPyPyPy-(R) <sup>β</sup> -H <sub>2</sub> N $\gamma$ -ImPyPyPy-(R) <sup>β</sup> -H <sub>2</sub> N $\gamma$ -) ( <b>1</b> ) .....	95
3.4.12 cyclo-(-ImPyPyPy-(R) <sup>β</sup> -AcHN $\gamma$ -ImPyPyPy-(R) <sup>β</sup> -H <sub>2</sub> N $\gamma$ -) ( <b>3</b> ) and cyclo-(-ImPyPyPy-(R) <sup>β</sup> -AcHN $\gamma$ -ImPyPyPy-(R) <sup>β</sup> -AcHN $\gamma$ -) ( <b>2</b> ) .....	96
3.4.13 ImPyPyPy-(R) <sup>β</sup> -H <sub>2</sub> N $\gamma$ -ImPyPyPy-(+)-IPA ( <b>4</b> ) .....	96
3.4.14 ImPyPyPy-(R) <sup>β</sup> -AcHN $\gamma$ -ImPyPyPy-(+)-IPA ( <b>5</b> ) .....	96
3.5 Notes and References .....	97
3.6 Spectra and Supplemental Information .....	99
Chapter 4: Oligomerization Route to Polyamide Macrocycles .....	117
Abstract .....	118
4.1 Introduction .....	119
4.2 Results and Discussion .....	120
4.3 Conclusion .....	121
4.4 Experimental Section .....	122
4.4.1 General .....	122
4.4.2 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> H .....	123
4.4.3 BocHN-(R) <sup>β</sup> -CbzHN $\gamma$ -ImPyPyPy-CO <sub>2</sub> Pfp ( <b>4</b> ) .....	123
4.4.4 Oligomerization procedure .....	123
4.4.5 UV Absorption Spectrophotometry .....	124
4.5 References and Notes .....	124
4.6 Spectra and Supplemental Information .....	126
Chapter 5: Allosteric Modulation of DNA by Small Molecules .....	127
Abstract .....	128
5.1 Introduction .....	129
5.2 Results and Discussion .....	130
5.3 Conclusion .....	134
5.4 Experimental Section .....	135
5.4.1 General .....	135
5.4.2 Synthesis .....	135
5.4.3 Cyclo-(-ImImPyPy-(R) <sup>α</sup> -BocHN $\gamma$ -ImImPyPy-(R) <sup>α</sup> -H <sub>2</sub> N $\gamma$ -) ( <b>6</b> ) .....	<b>135</b>
5.4.4 Cyclo-(-ImImPyPy-(R) <sup>α</sup> -H <sub>2</sub> N $\gamma$ -ImImPyPy-(R) <sup>α</sup> -H <sub>2</sub> N $\gamma$ -) ( <b>1</b> ) .....	<b>136</b>
5.4.5 Oligonucleotide Purification and Crystallization .....	137
5.4.6 UV-visible analysis .....	137
5.4.7 Data collection, Structure Determination, and Refinement .....	137
5.4.8 Structure Analysis and Figure Preparation .....	138
5.5 Notes and References .....	138
5.6 Spectra, Data Statistics, and Supplemental Information .....	141
Chapter 6: Structural Elucidation of a $\beta$ -amino- $\gamma$ -linked Cyclic Polyamide-DNA Complex and RNA Binding Studies .....	153

Abstract .....	154
6.1 Introduction .....	155
6.2 Overall Structure .....	157
6.3 Overall structure of DNA-polyamide complex .....	158
6.4 Turn conformation .....	159
6.5 Allosteric Perturbations .....	160
6.6 Solvation .....	161
6.7 RNA Binding Studies .....	162
6.8 Conclusion .....	163
6.9 Experimental .....	164
6.9.1 Synthesis .....	164
6.9.2 Oligonucleotide purification and Crystallization .....	164
6.9.3 Data collection, Structure determination, and refinement .....	164
6.9.4 Structure Analysis .....	165
6.10 Notes and References .....	165
6.11 Supplemental Information .....	168
Chapter 7: Programmable Oligomers Targeting 5'-GGGG-3' in the Minor Groove of DNA and NF- $\kappa$ B Binding Inhibition.....	171
Abstract .....	172
7.1 Introduction .....	173
7.2 Results and Discussion .....	175
7.2.1 Heterocycle Synthesis .....	175
7.2.2 Oligomer Synthesis .....	175
7.2.3 DNA affinity and sequence specificity .....	177
7.2.4 NF- $\kappa$ B electrophoretic mobility gel shift assay .....	178
7.3 Conclusion .....	183
7.4 Experimental .....	183
7.4.1 General .....	183
7.4.2 Heterocycle Synthesis .....	186
7.4.3 Oligomer Synthesis .....	189
7.4.4 Resin Cleavage Procedure .....	191
7.4.5 Footprinting Experiments .....	192
7.4.6 NF- $\kappa$ B Electrophoretic Mobility Shift Assay .....	192
7.5 Notes and References .....	194
Chapter 8: Fluorescent Sequence-Specific dsDNA Binding Oligomers .....	198
Abstract .....	199
8.1 Introduction .....	200
8.2 Results and Discussion .....	200
8.3 Conclusion .....	202
8.4 Experimental .....	202
8.5 Notes and References .....	203
8.6 Spectra and Supplemental Information .....	205
Chapter 9: Polyamide/NCP Ligation and Profluorescent Azido-Carbostyrils .....	209
Abstract .....	210
9.1 Introduction .....	211
9.1.1 Templated Dimerization of Polyamides .....	211
9.2 Results and Discussion .....	214
9.3 Conclusion .....	219
9.4 Experimental .....	220
9.4.1 Materials and General Methods .....	220

9.4.2 Plasmids .....	223
9.4.3 Polyamide Synthesis .....	223
9.4.4 Reconstitution of NCP .....	224
9.4.5 NCP Templated Ligation Reactions .....	224
9.5 Notes and References .....	224
Appendix A: Next Generation Hairpin Polyamides with (R)-3,4-Diaminobutyric Acid Turn Unit ...	226
Abstract .....	227
A.1 Introduction .....	228
A.2 Results and Discussion .....	228
A.2.1 Thermal stabilization of DNA duplexes by hairpin polyamides .....	228
A.2.2 Sequence apecificity at the turn position .....	232
A.2.3 Acetylated chiral hairpin polyamides.....	234
A.2.4 Biological assay for cell permeability.....	235
A.3 Conclusion .....	237
A.4 Experimental.....	237
A.4.1 General.....	237
A.4.2 Synthesis of polyamides .....	238
A.4.3 UV Absorption Spectrophotometry .....	239
A.4.4 Molecular Modeling.....	240
A.4.5 Measurement of Androgen-Induced PSA mRNA .....	240
A.5 Notes and References .....	240
A.6 Supplemental Information .....	243
Appendix B: Apredica ADMET Report .....	251
Appendix C: Programmable Oligomers for Minor Groove DNA Recognition .....	287
Abstract .....	288
C.1 Introduction.....	289
C.2 Experimental.....	290
C.2.1 Polyamide Synthesis .....	290
C.3 Results.....	292
C.3.1 DNA Affinity and Sequence Specificity of Dimer Caps .....	292
C.3.2 Design of a Programmable Oligomer for 5'-GTAC-3' .....	292
C.4 Discussion.....	293
C.5 Conclusion .....	297
C.6 Notes and References .....	297
C.7 Supplemental Information .....	299
C.7.1 General .....	299
C.7.2 Heterocycle Synthesis .....	300
C.7.3 Polyamide Synthesis .....	305
C.7.4 Deprotection of the O-Methyl-Protected Polyamides .....	306
C.7.5 Oligomer <b>9</b> Synthesis .....	307
C.7.6 O-Methyl Deprotection .....	309
C.7.7 Cleavage From Resin .....	309
Appendix D: Peptoid Cell Uptake Studies.....	310
Abstract .....	311
D.1 Summary.....	312
D.2 Experimental.....	315
D.2.1 Materials .....	315
D.2.2 General bromoacetic acid addition procedure .....	316
D.2.3 General amine addition procedure .....	316

D.2.4 Peptoid synthesis procedure (Synthesis of Resin <b>PR-1</b> ) .....	316
D.2.5 Procedure for first mini-PEG coupling (Synthesis of Resin <b>PR-2</b> ) .....	317
D.2.6 Preparation of peptoid 5-FAM ( <b>1</b> ) .....	318
D.2.7 Preparation of peptoid 6-FAM ( <b>2</b> ) .....	318
Appendix E: Cell Uptake Studies of 4G Targeting Polyamides .....	320
Abstract .....	321
E.1 Introduction .....	322
E.2 Results and Discussion .....	322
E.3 Conclusion .....	334
E.4 Experimental .....	335
E.5 Notes and References .....	335
E.6 Spectra and Supplemental Information .....	336

## List of Figures

### CHAPTER 1: Introduction to Molecular Recognition of DNA

<b>Figure 1.1</b> Chemical structure of DNA.....	4
<b>Figure 1.2</b> DNA base pairs.....	5
<b>Figure 1.3</b> DNA sugar phosphate backbone.....	6
<b>Figure 1.4</b> DNA polymorphs and A-form RNA.....	7
<b>Figure 1.5</b> Nucleosome core particle structures.....	9
<b>Figure 1.6</b> Anatomy of the DNA base pair edges.....	10
<b>Figure 1.7</b> X-ray structures of DNA binding transcription factors.....	11
<b>Figure 1.8</b> Atomic model of the interferon- $\beta$ enhancesome.....	12
<b>Figure 1.9</b> DNA-binding natural products.....	13
<b>Figure 1.10</b> DNA recognition by netropsin and distamycin.....	14
<b>Figure 1.11</b> Crystal structure of a 2:1 binding single strand Py/Im polyamide.....	15
<b>Figure 1.12</b> Crystal structure of a 2:1 binding single strand ImHpPyPy- $\beta$ -Dp polyamide...	16
<b>Figure 1.13</b> Polyamide pairing rules.....	17
<b>Figure 1.14</b> Consequence of covalent attachment of two polyamide strands.....	18
<b>Figure 1.15</b> Polyamide GABA-based turns.....	19
<b>Figure 1.16</b> NMR models of 1:1 and 6-ring cyclic polyamide DNA complexes.....	20
<b>Figure 1.17</b> Polyamide 2:1 DNA crystal structures colored by B-factor.....	21
<b>Figure 1.18</b> X-ray crystal structures of polyamide-NCP complexes.....	22
<b>Figure 1.19</b> Polyamide clamp bound to the NCP.....	22
<b>Figure 1.20</b> Current state of macromolecular crystallography: A DNA-drug perspective....	24

### CHAPTER 2: Solution-Phase Synthesis of Pyrrole–Imidazole Polyamides

<b>Figure 2.1</b> Structure of Py-Im hairpin polyamide <b>1</b> .....	33
<b>Figure 2.2</b> Retrosynthetic strategy for synthesis of polyamide <b>1</b> .....	34

<b>Figure 2.3</b> Analysis of polyamide <b>2</b> purity by analytical HPLC .....	37
<b>Figure 2.4</b> UV properties of polyamide <b>1</b> .....	38
<b>Figure 2.5</b> $^1\text{H}$ NMR of $\text{HCl}\cdot\text{H}_2\text{N-Py-CO}_2\text{Me}$ ( <b>10</b> ) .....	50
<b>Figure 2.6</b> $^{13}\text{C}$ NMR of $\text{HCl}\cdot\text{H}_2\text{N-Py-CO}_2\text{Me}$ ( <b>10</b> ) .....	51
<b>Figure 2.7</b> $^1\text{H}$ NMR of $\text{BocHN-PyPy-CO}_2\text{Me}$ ( <b>11</b> ) .....	52
<b>Figure 2.8</b> $^{13}\text{C}$ NMR of $\text{BocHN-PyPy-CO}_2\text{Me}$ ( <b>11</b> ) .....	53
<b>Figure 2.9</b> $^1\text{H}$ NMR of $\text{HCl}\cdot\text{H}_2\text{N-PyPy-CO}_2\text{Me}$ ( <b>12</b> ) .....	54
<b>Figure 2.10</b> $^{13}\text{C}$ NMR of $\text{HCl}\cdot\text{H}_2\text{N-PyPy-CO}_2\text{Me}$ ( <b>12</b> ) .....	55
<b>Figure 2.11</b> $^1\text{H}$ NMR of $\text{BocHN-PyPyPy-CO}_2\text{Me}$ ( <b>13</b> ) .....	56
<b>Figure 2.12</b> $^{13}\text{C}$ NMR of $\text{BocHN-PyPyPy-CO}_2\text{Me}$ ( <b>13</b> ) .....	57
<b>Figure 2.13</b> $^1\text{H}$ NMR of $\text{HCl}\cdot\text{H}_2\text{N-PyPyPy-CO}_2\text{Me}$ ( <b>4</b> ) .....	58
<b>Figure 2.14</b> $^{13}\text{C}$ NMR of $\text{HCl}\cdot\text{H}_2\text{N-PyPyPy-CO}_2\text{Me}$ ( <b>4</b> ) .....	59
<b>Figure 2.15</b> $^1\text{H}$ NMR of $\text{ImPyPyPy-CO}_2\text{Me}$ ( <b>14</b> ) .....	60
<b>Figure 2.16</b> $^{13}\text{C}$ NMR of $\text{ImPyPyPy-CO}_2\text{Me}$ ( <b>14</b> ) .....	61
<b>Figure 2.17</b> $^1\text{H}$ NMR of $\text{ImPyPyPy-CO}_2\text{H}$ ( <b>3</b> ) .....	62
<b>Figure 2.18</b> $^{13}\text{C}$ NMR of $\text{ImPyPyPy-CO}_2\text{H}$ ( <b>3</b> ) .....	63
<b>Figure 2.19</b> $^1\text{H}$ NMR of $\text{BocHN-}(R)^{\beta\text{-CbzHN}}\gamma\text{-Im-CO}_2\text{Et}$ ( <b>15</b> ) .....	64
<b>Figure 2.20</b> $^{13}\text{C}$ NMR of $\text{BocHN-}(R)^{\beta\text{-CbzHN}}\gamma\text{-Im-CO}_2\text{Et}$ ( <b>15</b> ) .....	65
<b>Figure 2.21</b> $^1\text{H}$ NMR of $\text{HCl}\cdot\text{H}_2\text{N-}(R)^{\beta\text{-CbzHN}}\gamma\text{-Im-CO}_2\text{Et}$ ( <b>5</b> ) .....	66
<b>Figure 2.22</b> $^{13}\text{C}$ NMR of $\text{HCl}\cdot\text{H}_2\text{N-}(R)^{\beta\text{-CbzHN}}\gamma\text{-Im-CO}_2\text{Et}$ ( <b>5</b> ) .....	67
<b>Figure 2.23</b> $^1\text{H}$ NMR of $\text{ImPyPyPy-}(R)^{\beta\text{-CbzHN}}\gamma\text{-Im-CO}_2\text{Et}$ ( <b>16</b> ) .....	68
<b>Figure 2.24</b> $^{13}\text{C}$ NMR of $\text{ImPyPyPy-}(R)^{\beta\text{-CbzHN}}\gamma\text{-Im-CO}_2\text{Et}$ ( <b>16</b> ) .....	69
<b>Figure 2.25</b> $^1\text{H}$ NMR of $\text{ImPyPyPy-}(R)^{\beta\text{-CbzHN}}\gamma\text{-Im-CO}_2\text{H}$ ( <b>17</b> ) .....	70
<b>Figure 2.26</b> $^{13}\text{C}$ NMR of $\text{ImPyPyPy-}(R)^{\beta\text{-CbzHN}}\gamma\text{-Im-CO}_2\text{H}$ ( <b>17</b> ) .....	71
<b>Figure 2.27</b> $^1\text{H}$ NMR of $\text{ImPyPyPy-}(R)^{\beta\text{-CbzHN}}\gamma\text{-ImPyPyPy-CO}_2\text{Me}$ ( <b>2</b> ) .....	72

<b>Figure 2.28</b> $^{13}\text{C}$ NMR of ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>2</b> ) .....	73
<b>Figure 2.29</b> $^1\text{H}$ NMR of ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> H ( <b>22</b> ).....	74
<b>Figure 2.30</b> $^{13}\text{C}$ NMR of ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> H ( <b>22</b> ).....	75
<b>Figure 2.31</b> $^1\text{H}$ NMR of BocHN-(+)- <sup>BnO</sup> IPA ( <b>20</b> ).....	76
<b>Figure 2.32</b> $^{13}\text{C}$ NMR of BocHN-(+)- <sup>BnO</sup> IPA ( <b>20</b> ).....	77
<b>Figure 2.33</b> $^1\text{H}$ NMR of ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-(+)- <sup>BnO</sup> IPA ( <b>23</b> ).....	78
<b>Figure 2.34</b> $^{13}\text{C}$ NMR of ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-(+)- <sup>BnO</sup> IPA ( <b>23</b> ).....	79
<b>Figure 2.35</b> $^1\text{H}$ NMR of ImPyPyPy-( <i>R</i> ) $^{\beta\text{-H}2\text{N}\gamma}$ -ImPyPyPy-(+)-IPA ( <b>1</b> ).....	80
<b>Figure 2.36</b> $^{13}\text{C}$ NMR of ImPyPyPy-( <i>R</i> ) $^{\beta\text{-H}2\text{N}\gamma}$ -ImPyPyPy-(+)-IPA ( <b>1</b> ).....	81

### CHAPTER 3: Cyclic Pyrrole–Imidazole Polyamides Targeted to the Androgen Response Element

<b>Figure 3.1</b> Structures of cyclic and hairpin polyamides <b>1–5</b> .....	84
<b>Figure 3.2</b> Targeting the ARE with DNA-binding polyamides.....	88
<b>Figure 3.3</b> Polyamides <b>1</b> and <b>5</b> ADMET testing .....	99
<b>Figure 3.4</b> $^1\text{H}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -Im-CO <sub>2</sub> H ( <b>8</b> ).....	101
<b>Figure 3.5</b> $^{13}\text{C}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -Im-CO <sub>2</sub> H ( <b>8</b> ).....	102
<b>Figure 3.6</b> $^1\text{H}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>9</b> ).....	103
<b>Figure 3.7</b> $^{13}\text{C}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>9</b> ).....	104
<b>Figure 3.8</b> $^1\text{H}$ NMR HCl•H <sub>2</sub> N-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>10</b> ).....	105
<b>Figure 3.9</b> $^{13}\text{C}$ NMR HCl•H <sub>2</sub> N-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>10</b> ).....	106
<b>Figure 3.10</b> $^1\text{H}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> H ( <b>11</b> ).....	107
<b>Figure 3.11</b> $^{13}\text{C}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> H ( <b>11</b> ).....	108
<b>Figure 3.12</b> $^1\text{H}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>12</b> ) .	109
<b>Figure 3.13</b> $^{13}\text{C}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Me ( <b>12</b> )	110
<b>Figure 3.14</b> $^1\text{H}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> H ( <b>13</b> )....	111

<b>Figure 3.15</b> $^{13}\text{C}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> H ( <b>13</b> ) ...	112
<b>Figure 3.16</b> $^1\text{H}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Pfp ( <b>14</b> ) .	113
<b>Figure 3.17</b> $^1\text{H}$ NMR <i>cyclo</i> -(ImPyPyPy-( <i>R</i> ) $^{\beta\text{-H}2\text{N}\gamma}$ -ImPyPyPy-( <i>R</i> ) $^{\beta\text{-H}2\text{N}\gamma}$ -) ( <b>1</b> ) .....	114
<b>Figure 3.18</b> Analytical HPLC characterization of cyclic polyamide <b>15</b> .....	115
<b>Figure 3.19</b> Analytical HPLC characterization of cyclic polyamide <b>1</b> .....	115
<b>Figure 3.20</b> Analytical HPLC characterization of cyclic polyamide <b>3</b> .....	115
<b>Figure 3.21</b> Analytical HPLC characterization of cyclic polyamide <b>2</b> .....	116
<b>Figure 3.22</b> Analytical HPLC characterization of cyclic polyamide <b>5</b> .....	116

#### CHAPTER 4: Oligomerization Route to Polyamide Macrocycles

<b>Figure 4.1</b> Structures of macrocyclic polyamides <b>1z–3z</b> and <b>1–3</b> .....	119
<b>Figure 4.2</b> Synthesis of macrocyclic polyamides <b>1–3</b> .....	120
<b>Figure 4.3</b> Reverse phase HPLC analysis of the oligomerization reaction.....	121
<b>Figure 4.4</b> $^1\text{H}$ NMR BocHN-( <i>R</i> ) $^{\beta\text{-CbzHN}\gamma}$ -ImPyPyPy-CO <sub>2</sub> Pfp ( <b>4</b> ) .....	126

#### CHAPTER 5: Allosteric Modulation of DNA by Small Molecules

<b>Figure 5.1</b> Chemical structure of the cyclic polyamide <b>1</b> .....	129
<b>Figure 5.2</b> Comparison of native DNA to polyamide/DNA complex.....	130
<b>Figure 5.3</b> Analysis of groove width for native DNA and polyamide complexed DNA .....	131
<b>Figure 5.4</b> Conformation of the $\alpha$ -amino substituted GABA turn .....	132
<b>Figure 5.5</b> Direct and water-mediated non-covalent molecular recognition interactions....	134
<b>Figure 5.6</b> Solid-phase synthesis of cyclic polyamide <b>1</b> .....	141
<b>Figure 5.7</b> Polyamide analytical data.....	142
<b>Figure 5.8</b> Single crystal of the polyamide-DNA complex and UV-Vis.....	144
<b>Figure 5.9</b> Hydrogen bond map of polyamide-DNA complex.....	145
<b>Figure 5.10</b> Polyamide/DNA complex map.....	146

<b>Figure 5.11</b> DNA structure map .....	147
<b>Figure 5.12</b> Hydrogen bond map of DNA crystal structure .....	148
<b>Figure 5.13</b> Molecular electrostatic potential maps of compounds <b>7-10</b> and complex <b>11</b> ..	149
<b>Figure 5.14</b> Comparison of Local base-pair step parameters .....	150
<b>Figure 5.15</b> Comparison of Local base-pair parameters .....	151
<b>Figure 5.16</b> Comparison of Local base-pair helical parameters .....	152
 CHAPTER 6: Structural Elucidation of a $\beta$ -amino- $\gamma$ -linked Cyclic Polyamide-DNA Complex and RNA Binding Studies	
<b>Figure 6.1</b> Structure of cyclic polyamide–DNA complex at 0.95 Å resolution.....	156
<b>Figure 6.2</b> Molecular recognition details from the X-ray structure .....	158
<b>Figure 6.3</b> DNA minor and major groove dimensions.....	159
<b>Figure 6.4</b> Allosteric distortion upon polyamide binding .....	160
<b>Figure 6.5</b> Polyamide–RNA binding.....	162
<b>Figure 6.6</b> Comparison of Local base-pair step parameters.....	168
<b>Figure 6.7</b> Comparison of Local base-pair parameters .....	169
<b>Figure 6.8</b> Comparison of Local base-pair helical parameters .....	170
 CHAPTER 7: Programmable Oligomers Targeting 5'-GGGG-3' in the Minor Groove of DNA and NF- $\kappa$ B Binding Inhibition	
<b>Figure 7.1</b> Structures of pyrrole-benzimidazole dimers .....	174
<b>Figure 7.2</b> Crystal structure of the NF- $\kappa$ B heterodimer bound to the DNA .....	174
<b>Figure 7.3</b> Postulated hydrogen-bonding models for the polyamide–DNA complexes .....	176
<b>Figure 7.4</b> Illustration of NF- $\kappa$ B:DNA binding inhibition by oligomer <b>2</b> .....	177
<b>Figure 7.5</b> DNaseI footprinting for polyamides <b>1</b> and <b>2</b> .....	180
<b>Figure 7.6</b> DNaseI footprinting for polyamides <b>3</b> and <b>4</b> .....	181
<b>Figure 7.7</b> DNaseI footprinting for polyamides <b>5</b> and <b>6</b> .....	182

<b>Figure 7.8</b> Gel shift screen for compounds <b>1–6</b> .....	184
<b>Figure 7.9</b> EC <sub>50</sub> value for compound <b>2</b> from gel shift experiment .....	185
<b>Figure 7.10</b> Antibody supershift on match DNA with anti-p50 or anti-p65 antibody .....	185
CHAPTER 8: Fluorescent Sequence-Specific dsDNA Binding Oligomers	
<b>Figure 8.1</b> Structure of oligomers <b>O1</b> and <b>O2</b> .....	200
<b>Figure 8.2</b> Design of dsDNA library .....	201
<b>Figure 8.3</b> Fluorescence emission spectra of <b>O1</b> and <b>O2</b> .....	202
<b>Figure 8.4</b> Plot of dsDNA concentration versus fluorescence .....	203
<b>Figure 8.5</b> UV spectra for compounds <b>O1</b> and <b>O2</b> .....	205
<b>Figure 8.6</b> Fluorescence emission as a function of DNA concentration .....	206
<b>Figure 8.7</b> Fluorescence emission spectra compound <b>O1</b> .....	207
<b>Figure 8.8</b> Fluorescence emission spectra compound <b>O2</b> .....	208
CHAPTER 9: Polyamide/NCP Ligation and Profluorescent Azido-Carbostyrils	
<b>Figure 9.1</b> Crystal structure of an NCP bound polyamide clamp .....	211
<b>Figure 9.2</b> Illustration of the clamp dimerization strategy .....	212
<b>Figure 9.3</b> Sequence of the 146 base pair fragment of $\alpha$ -satellite DNA .....	213
<b>Figure 9.4</b> Analysis of the linker dependence .....	214
<b>Figure 9.5</b> Synthesis of alkyl azido linkers .....	214
<b>Figure 9.6</b> Synthesis of azido-polyamides .....	215
<b>Figure 9.7</b> Synthesis of alkynyl-polyamides .....	216
<b>Figure 9.8</b> DNA templated ligation on the NCP .....	217
<b>Figure 9.9</b> Control reactions for the NCP templated ligation .....	218
<b>Figure 9.10</b> Linker distances for incorporation of the profluorescent coumarin .....	219
<b>Figure 9.11</b> Carbostyril linker distance .....	220

<b>Figure 9.12</b> Synthesis of profluorescent azide <b>17</b> .....	221
<b>Figure 9.13</b> Photophysical properties of compound <b>17</b> .....	222
<b>Figure 9.14</b> Representative gels of the NCP reconstitution .....	224

#### APPENDIX A: Next Generation Hairpin Polyamides with (R)-3,4-Diaminobutyric Acid Turn Unit

<b>Figure A.1</b> Increased DNA-binding affinity caused by different $\gamma$ -turn units .....	229
<b>Figure A.2</b> Chemical structures for hairpins <b>1-16</b> .....	230
<b>Figure A.3</b> Normalized UV denaturation profiles.....	231
<b>Figure A.4</b> Models of different turn conformations .....	233
<b>Figure A.5</b> Chemical structures polyamides <b>17-19</b> .....	234
<b>Figure A.6</b> Androgen receptor (AR)-mediated transcription .....	235
<b>Figure A.7</b> Chemical structures polyamides <b>20-23</b> .....	236
<b>Figure A.8</b> DNase I footprint titration experiments (polyamides <b>2-4</b> ) .....	245
<b>Figure A.9</b> DNase I footprint titration experiments (polyamides <b>6-8</b> ) .....	246
<b>Figure A.10</b> DNase I footprint titration experiments (polyamides <b>10-12</b> ).....	247
<b>Figure A.11</b> DNase I footprint titration experiments (polyamides <b>14-16</b> ) .....	248
<b>Figure A.12</b> A) Chemical structures of polyamides <b>20-23</b> .....	250

#### APPENDIX B: Apredica ADMET Report

<b>Figure B.1</b> Polyamides <b>1</b> and <b>5</b> .....	252
--	-----

#### APPENDIX C: Programmable Oligomers for Minor Groove DNA Recognition

<b>Figure C.1</b> Structures of dimers.....	290
<b>Figure C.2</b> Postulated hydrogen-bonding models for polyamide-DNA complexes.....	291
<b>Figure C.3</b> Quantitative DNase I footprinting experiments for polyamides <b>1, 2, and 3</b> .....	294
<b>Figure C.4</b> Postulated hydrogen-bonding model and structure of oligomer <b>9</b> .....	296

## APPENDIX D: Peptoid Cell Uptake Studies

<b>Figure D.1</b> Compound <b>1</b> , 5-FAM, and <b>2</b> , 6-FAM. ....	312
<b>Figure D.2</b> Synthesis of compound <b>1</b> and <b>2</b> . ....	313
<b>Figure D.3</b> HeLa cell uptake studies for compound <b>1</b> .....	314
<b>Figure D.4</b> HeLa cell uptake studies for compound <b>2</b> .....	315

## APPENDIX E: Cell Uptake Studies of 4G Targeting Polyamides

<b>Figure E.1</b> Fluorescent polyamide <b>2</b> .....	322
<b>Figure E.2</b> Tail and turn modifications for 23 compound polyamide library. ....	323
<b>Figure E.3</b> Fluorescent polyamide cell uptake studies (Compounds <b>1-4</b> ) .....	326
<b>Figure E.4</b> Fluorescent polyamide cell uptake studies (Compounds <b>5-8</b> ) .....	327
<b>Figure E.5</b> Fluorescent polyamide cell uptake studies (Compounds <b>9-11</b> ) .....	327
<b>Figure E.6</b> Fluorescent polyamide cell uptake studies (Compounds <b>12-15</b> ) .....	328
<b>Figure E.7</b> Fluorescent polyamide cell uptake studies (Compounds <b>16-19</b> ) .....	329
<b>Figure E.8</b> Fluorescent polyamide cell uptake studies (Compounds <b>20-23</b> ) .....	330
<b>Figure E.9</b> Mechanism of mitochondrial staining using MitoTracker.....	331
<b>Figure E.10</b> Results of HeLa cell uptake study using MitoTracker and compound <b>4</b> .....	331
<b>Figure E.11</b> Results of HeLa cell uptake study using LysoTracker and compound <b>4</b> .....	332
<b>Figure E.12</b> Cell uptake with ( $\pm$ )-verapamil (Compound <b>23</b> ) .....	333
<b>Figure E.13</b> Cell uptake study with Lipofectamine 2000 (Compound <b>23</b> ) .....	334

## List of Schemes

### CHAPTER 2: Solution-Phase Synthesis of Pyrrole–Imidazole Polyamides

<b>Scheme 2.1</b> Preparation of <b>3</b> and <b>4</b> .....	35
<b>Scheme 2.2</b> Preparation of <b>5</b> and assembly of core polyamide <b>2</b> .....	36
<b>Scheme 2.3</b> Preparation of <b>21</b> .....	37
<b>Scheme 2.4</b> Final steps for the synthesis of Py-Im polyamide <b>1</b> .....	37

### CHAPTER 3: Cyclic Pyrrole–Imidazole Polyamides Targeted to the Androgen Response Element

<b>Scheme 3.1</b> Preparation of <b>10</b> and <b>11</b> .....	85
<b>Scheme 3.2</b> Preparation of <b>1</b> , <b>2</b> , and <b>3</b> .....	86

### CHAPTER 7: Programmable Oligomers Targeting 5'-GGGG-3' in the Minor Groove of DNA and NF- $\kappa$ B Binding Inhibition

<b>Scheme 7.1.</b> Synthesis of imidazopyridine–imidazole dimers.....	178
<b>Scheme 7.2.</b> Representative solid-phase synthesis of polyamide <b>6</b> . .....	179

### APPENDIX C: Programmable Oligomers for Minor Groove DNA Recognition

<b>Scheme C.1</b> Representative solid-phase synthesis of polyamide <b>2</b> and <b>3</b> .....	293
<b>Scheme C.2</b> Synthesis of <b>15</b> .....	301
<b>Scheme C.3</b> Synthesis of <b>7</b> .....	302
<b>Scheme C.4</b> Synthesis of <b>8</b> .....	303
<b>Scheme C.5</b> Synthesis of <b>1</b> .....	304
<b>Scheme C.6</b> Compound <b>2</b> . .....	305
<b>Scheme C.7</b> Compound <b>3</b> . .....	305
<b>Scheme C.8</b> Synthesis of <b>9-13</b> . .....	307

### APPENDIX E: Cell Uptake Studies of 4G Targeting Polyamides

<b>Scheme E.1.</b> Synthesis of <b>R4</b> .....	324
<b>Scheme E.2.</b> Synthesis of compound library .....	325

## List of Tables

CHAPTER 1: Introduction to Molecular Recognition of DNA	
<b>Table 1.1</b> Typical nucleic acid structural parameters. ....	8
CHAPTER 3: Cyclic Pyrrole–Imidazole Polyamides Targeted to the Androgen Response Element	
<b>Table 3.1</b> $T_m$ values for polyamides for <b>1–5</b> .....	87
<b>Table 3.2</b> Caco-2 permeability summary.....	99
<b>Table 3.3</b> Cytotoxicity summary .....	99
<b>Table 3.4</b> Fluorescent Cyp IC50 summary .....	100
<b>Table 3.5</b> hERG FastPatch summary .....	100
<b>Table 3.6</b> Plasma half-life summary .....	100
<b>Table 3.7</b> Plasma protein binding summary .....	100
CHAPTER 4: Oligomerization Route to Polyamide Macrocycles	
<b>Table 4.1</b> $T_m$ values for cycles 1–3 in the presence of DNA. <sup>a</sup> .....	122
CHAPTER 5: Allosteric Modulation of DNA by Small Molecules	
<b>Table 5.1</b> Data collection and refinement statistics. ....	143
CHAPTER 6: Structural Elucidation of a $\beta$ -amino- $\gamma$ -linked Cyclic Polyamide-DNA Complex and RNA Binding Studies	
<b>Table 6.1</b> Data collection and refinement statistics. ....	157
<b>Table 6.2</b> Buckle and opening values.....	161
<b>Table 6.3</b> Polyamide-DNA and Polyamide-RNA melting temperatures.....	162
CHAPTER 7: Programmable Oligomers Targeting 5'-GGGG-3' in the Minor Groove of DNA and NF- $\kappa$ B Binding Inhibition	
<b>Table 7.1</b> Affinities of 5'-GGGG-3' binding oligomers. ....	183

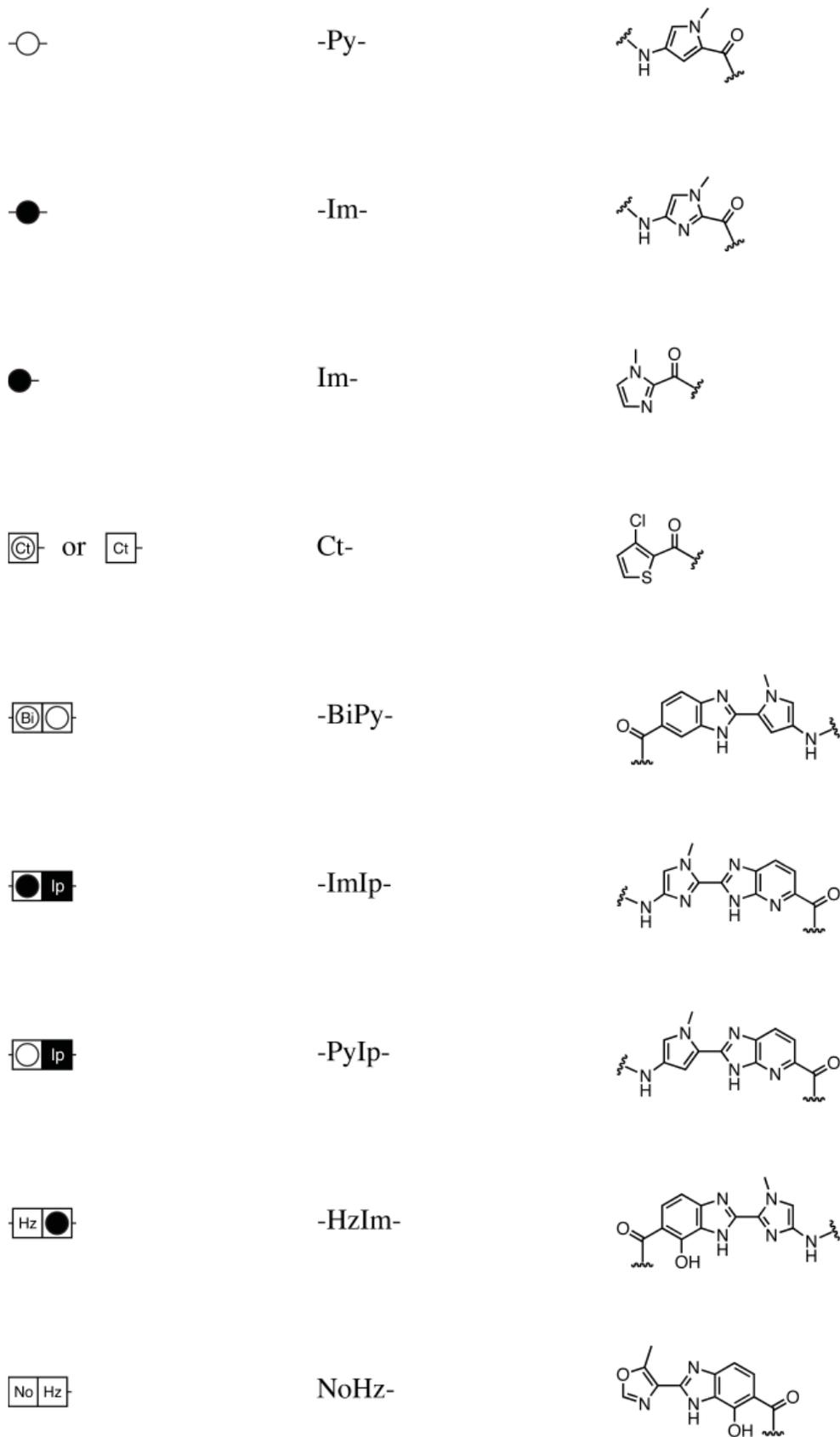
## APPENDIX A: Next Generation Hairpin Polyamides with (R)-3,4-Diaminobutyric Acid Turn Unit

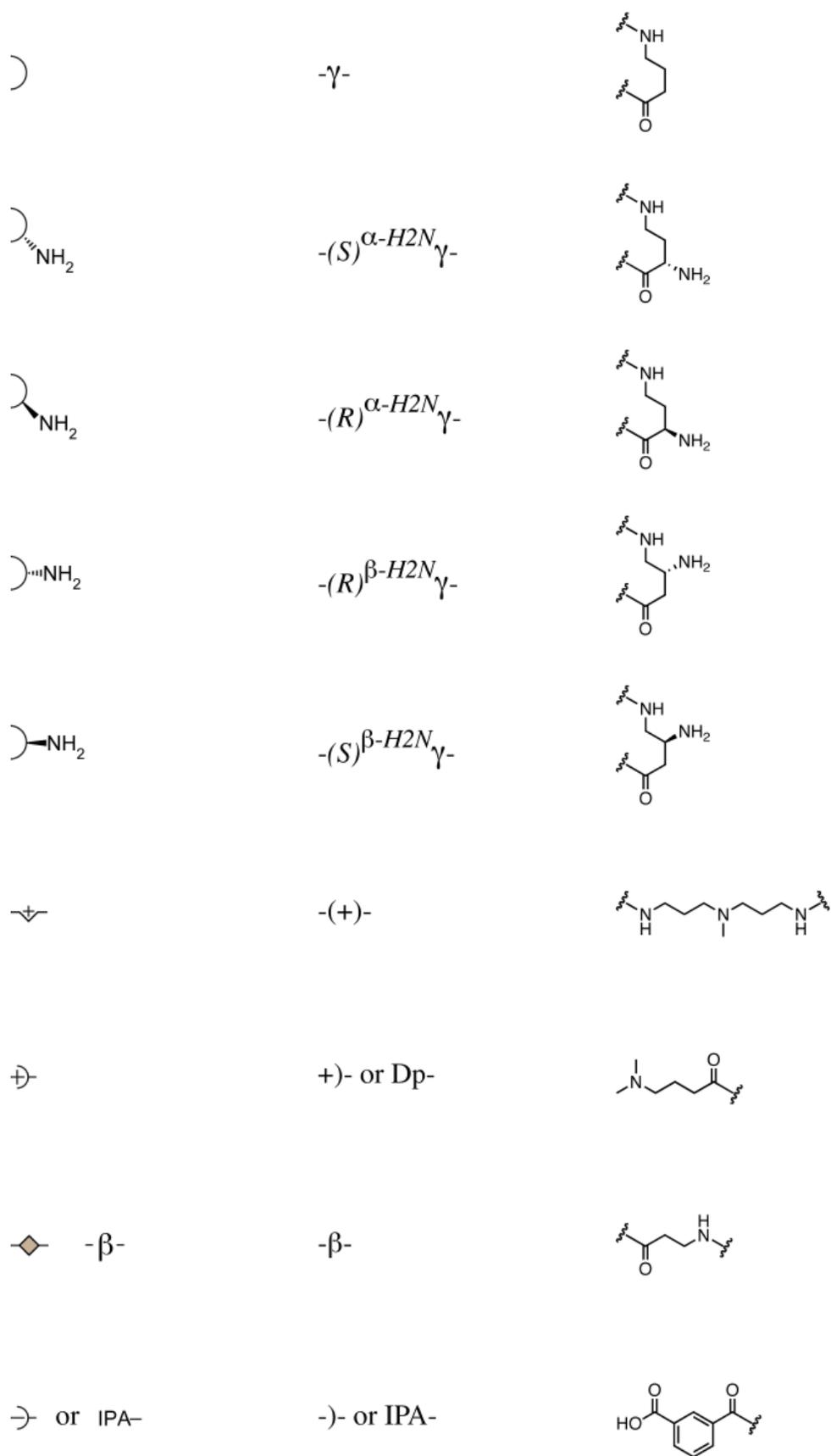
<b>Table A.1</b> Melting temperatures of DNA/polyamide complexes .....	232
<b>Table A.2</b> Melting temperatures of DNA/polyamide complexes .....	235
<b>Table A.3</b> Melting temperatures of DNA/polyamide complexes .....	243
<b>Table A.4</b> Equilibrium association constants for hairpin polyamides .....	244
<b>Table A.5</b> Melting temperatures of polyamides .....	249
Appendix C: Programmable Oligomers for Minor Groove DNA Recognition .....	287
<b>Table C.1</b> Affinities of X/Py ring pairs .....	295

## Nomenclature and Symbology

A	adenine
Å	angstrom
A•T	adenine Watson-Crick hydrogen bonded to thymine
Ac <sub>2</sub> O	acetic anhydride
ADMET	absorption, distribution, metabolism, excretion, and toxicity
AP	activating protein
AR	androgen receptor
ARE	androgen response element
atm	atmosphere
Bi	benzimidazole
Boc	<i>tert</i> -butyloxycarbonyl
bp	base pair
°C	degrees Celsius
C	cytosine
calc'd	calculated
Cbz	carbobenzyloxy
CCDC	Cambridge Crystallographic Data Centre
Ct	2-carboxy-3-chlorothiophene
Dbu	diazabicycloundecane
DCM	dichloromethane
DHT	dihydrotestosterone
DIEA	<i>N,N</i> -diisopropylethylamine
DMF	<i>N,N</i> -dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
Dp	<i>N,N</i> -dimethylaminopropylamine
ds	double strand
Em	emission
ESI	electrospray ionization
Et	ethyl
Ex	excitation
FAB	fast-atom bombardment
Fmoc	fluorenylmethyloxycarbonyl
G	guanine
g	grams
G•C	guanine Watson-Crick hydrogen bonded to cytosine
GABA	gamma-aminobutyric acid
h	hour(s)
HBTU	2-(1 <i>H</i> -benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HF	hartree fock
Hp	3-hydroxypyrrole
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry

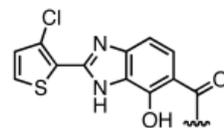
h $\nu$	light
Hz	hydroxybenzimidazole
IC <sub>50</sub>	median inhibition concentration (50%)
Im	<i>N</i> -methylimidazole
Ip	imidazopyridine
K <sub>a</sub>	association constant
K <sub>d</sub>	dissociation constant
$\lambda$	wavelength
<i>m/z</i>	mass to charge ratio
$\mu$	micro
M	molar
max	maximum
MALDI	Matrix-assisted LASER desorption/ionization
min	minute(s)
mol	mole(s)
mmol	millimole(s)
MS	mass spectrometry
N	normal
N	A, T, G, or C
NCP	nucleosome core particle
No	oxazole
NOESY	nuclear Overhauser enhancement spectroscopy
PCR	polymerase chain reaction
Py-Im	pyrrole-imidazole
PNA	peptide nucleic acid
PSA	prostate specific antigen
Py-Im	pyrrole-imidazole
RT-PCR	reverse transcriptase PCR
Py	<i>N</i> -methylpyrrole
PyBOP	(benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
OBt	hydroxytriazole ester
R <sub>f</sub>	retention factor
RNA	ribonucleic acid
RP-HPLC	reverse-phase high performance liquid chromatography
sat.	saturated
satd.	saturated
ss	single strand
T	thymine
TFA	trifluoroacetic acid
TMR	tetramethyl rhodamine
TO	thiazole orange
TOF	time-of-flight
U	uracil
UV	ultraviolet
Vis	visible
W	adenine or thymine



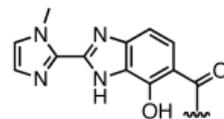




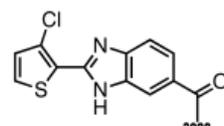
CtHz-



ImHz-



CtBi-



## **Chapter 1: Introduction to Molecular Recognition of DNA**

## 1.1 Background and Significance

The field of molecular recognition has come a long way since the organic solvent based host-guest chemistry of Lehn and Cram (crown ether cation complexes).<sup>1-3</sup> Understanding in a predictive and mechanistic sense the molecular recognition between synthetic ligands and biological macromolecules in water is fundamental to understanding biochemical processes and cellular composition.<sup>4</sup> The overall free energy of these complexes often includes a superposition of non-covalent forces such as hydrogen bonding interactions, dipole-dipole, induced dipole, cation- $\pi$ , lone pair- $\pi$ , and van der Waals interactions in addition to hydrophobic effects. Understanding the intimate interplay of these forces and their contributions to the overall free energy of a host-guest system has remained one of the ultimate challenges in chemistry and biology. The molecular recognition processes involved in nucleic acid-drug and nucleic acid-protein interactions are similar with both being driven by the hydrophobic effect, a phenomenon which is still not well understood. As ligand-receptor recognition proceeds, the optimization of multiple forces ensues including minimization of water exposed hydrophobic surfaces and simultaneous maximization of van der Waals interactions. Additionally, optimization of all buried hydrogen-bond donor and acceptor pairings including solvent-assisted and counterion charge neutralization contribute to the complex recognition event. Intimate structural and biophysical knowledge of these processes is fundamental to the understanding of nature at the molecular level.

The DNA double helix, in addition to being the molecular storage unit of genetic information, represents one of the ultimate challenges in aqueous based molecular recognition. Over billions of years, nature has used selection to evolve protein surfaces that recognize DNA in a cooperative and combinatorial fashion allowing for the stringent regulation of the molecular processes crucial to all living organisms on earth. Prior to the 1960s, histologists and cell biologists realized that certain small molecules could interact specifically with cell nuclei.<sup>5</sup> Dye molecules such as aminoacridines were regularly used for staining tissues and cells and it was recognized that specificity for different nucleic acid structures could be obtained using different dyes. However it was not until the 1960s that a formal DNA drug binding hypothesis would be formulated. The “intercalation hypothesis” formulated by Leonard Lerman (a graduate student of Linus Pauling at Caltech) in 1961, working at the Cambridge MRC laboratory, provided the pivotal turning point in the field of drug-nucleic acid recognition.<sup>6</sup> Since the intercalation hypothesis, a plethora of biophysical, biochemical, and biostructural investigations have unveiled the detailed chemistry and biology of many DNA binding drugs, some of which have had a profound impact on human disease (i.e. actinomycin D).<sup>5,7,8</sup> The

intercalating natural product actinomycin D remained one of the most potent chemotherapeutics throughout the 1950s and 1960s along with other nucleic acid binding drugs including cross-linking agents and powerful alkylators, however the first minor groove binding agents would not be discovered until the mid 1960s.<sup>5</sup> Even though Lerman himself relied upon X-ray fiber diffraction data for his intercalation hypothesis in the 1960s, it took another 15-20 years before the first single crystal X-ray structures of drug-nucleic acid complexes (intercalators) would emerge with the seminal work of Sobell, Rich, and Neidle.<sup>9-11</sup> The first X-ray structure of a minor groove binder would not appear until Dickerson's report on the 1:1 structure of netropsin complexed with DNA in 1985.<sup>12</sup> This was soon followed by the 1:1 structure of the distamycin-DNA complex by Rich in 1987.<sup>13</sup> In a seminal study, structural evidence using NMR for the 2:1 binding motif of distamycin was provided by Wemmer in 1989,<sup>14</sup> however the first single crystal X-ray structure of a 2:1 minor groove binding ligand-DNA complex was not realized until the work of Ramakrishnan in 1994 on distamycin A.<sup>15</sup> Since this work many advances have been made in the field of DNA molecular recognition, with minor groove binders representing one of the most promising classes of DNA-binding molecules for targeted transcriptional therapy.

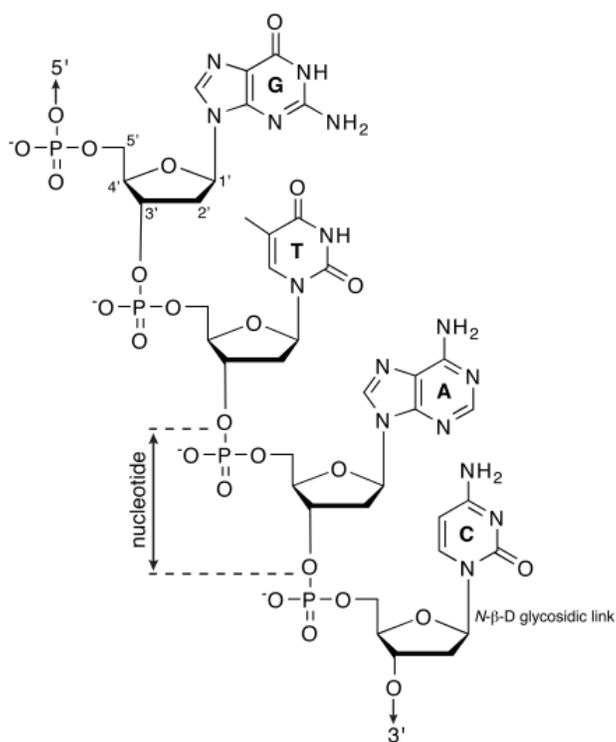
The modulation of gene expression using small molecules has been one of the ultimate goals of nucleic acid molecular recognition. Complex natural products such as actinomycin D, netropsin, and distamycin A have served as inspiration to chemists for the construction of molecular architectures capable of nucleic acid recognition with specificity and affinities equivalent to and rivaling that of endogenous proteins. Initially inspired by the 2:1 binding natural product distamycin, pyrrole-imidazole polyamides have evolved into a modular programmable molecular recognition system capable of specificities and affinities rivaling that of endogenous transcription factors.<sup>16,17</sup> Modulation of transcription factor-DNA interfaces with small molecules such as pyrrole-imidazole polyamides provides a powerful strategy for controlling regulation of the genetic material and could eventually impact human medicine. The future of molecular recognition is poised to benefit greatly from advances in biochemical, biophysical, computational, and structural (X-ray, NMR, EM, Cryo-EM, etc.) methods along with the new tools of physical biology leading to ever increasing resolution and a quantitative understanding of molecular level processes.<sup>4,18</sup>

## 1.2 Nucleic Acid Structure

Deoxyribose nucleic acid (DNA) is the fundamental storage material of genetic information and can be characterized chemically as a hetero-polymer consisting of nucleotide monomers linked

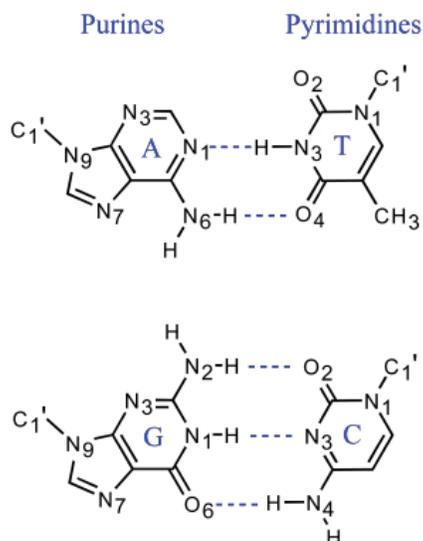
through their sugar-phosphate backbones.<sup>19,20</sup> The 5' and 3' hydroxyl groups of the deoxyribose sugar define the directionality of the DNA strand while a set of four nucleobases [adenine (A), guanine (G), cytosine (C), and thymine (T)] distinguish the nucleotide monomers, providing the fundamental building blocks of the genetic code. Figure 1.1 shows the chemical structure of a short DNA strand containing all four bases and Figure 1.2 shows the atom numbering conventions. Early studies by Chargaff demonstrated that A and T occurred in similar molar ratios as did G and C, which in combination with fiber diffraction data from Rosalind Franklin and Maurice Wilkins would eventually lead to Watson and Crick's base-paired helical model of B-DNA.<sup>21</sup> The Watson-Crick base paired model of DNA contains a set of rules for which A prefers to bind T through two hydrogen bonds and G prefers to bind C through three hydrogen bonds on opposite strands as shown in Figure 1.2 (U replaces T in RNA).<sup>22</sup> The strands are oriented in an antiparallel fashion as they base pair and wind around a central axis. These opposite strands form a double helical structure where the Watson-Crick base pairs are stacked and stabilized by a combination of favorable hydrophobic effects and hydrogen bonding between paired bases. Due to the length of the sugar-phosphate backbone, a helical twist is required to minimize the distance between adjacent base pairs and maximize their hydrophobic stacking.<sup>19,20</sup>

The sugar-phosphate backbone of DNA is highly dynamic allowing for a diverse range



**Figure 1.1** Chemical structure of DNA.

of higher order structures depending on environmental conditions. The torsion angles for the sugar-phosphate backbone are defined in Figure 1.3 and typically vary with ionic strength, pH, sequence, and many other factors.<sup>19,20</sup> In contrast to RNA, where the 2'-hydroxyl of the sugar locks the A-form helix into a fairly rigid structure, the sugar-phosphate backbone of DNA is highly mobile.<sup>19,20</sup> DNA conformation can often be defined by the sugar pucker modes, which by convention are named after the ring atom and either *endo* or *exo* referring to the 5' side of the furanose ring or the 3' side, respectively. Figure 1.3 shows typical



**Figure 1.2** DNA base pairs showing numbering convention for heteroatoms and Watson-Crick base pairing.

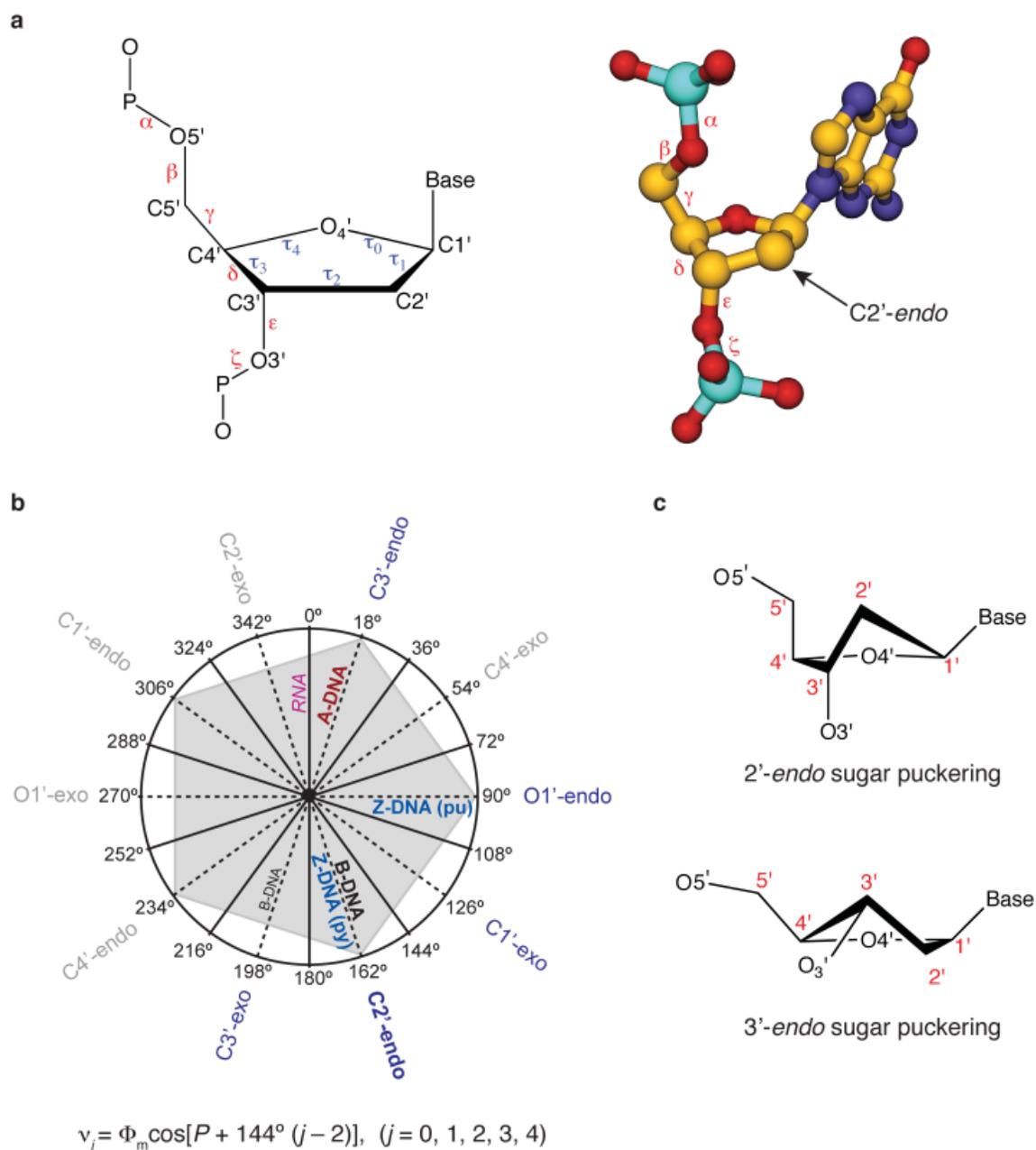
sugar pucker conformations along with their preference in nucleic acid structures. In addition to the sugars displaying conformational preferences, the phosphodiester bond exhibits conformational rigidity analogous to a peptide bond. This conformation rigidity, known as the gauche effect, is a result of stereoelectronic effects from lone pair hyperconjugation/donation of the O3' and O5' oxygen atoms into the  $\sigma^*$  orbital of the P-O5' and P-O3' bonds, respectively.<sup>23</sup> Double helical DNA is a dynamic structure which is capable of forming three primary double strand conformations known as A, B, and Z forms. In contrast to this, double helical RNA is far less flexible with the 2'-OH locking its sugar ring conformation into a C3'-endo pucker resulting in a preference for an A form helix similar to that of A-form DNA. A structural comparison of these ideal DNA polymorphs along with A-form RNA is

shown in Figure 1.4 and Table 1.1.<sup>19,20,23,24.</sup>

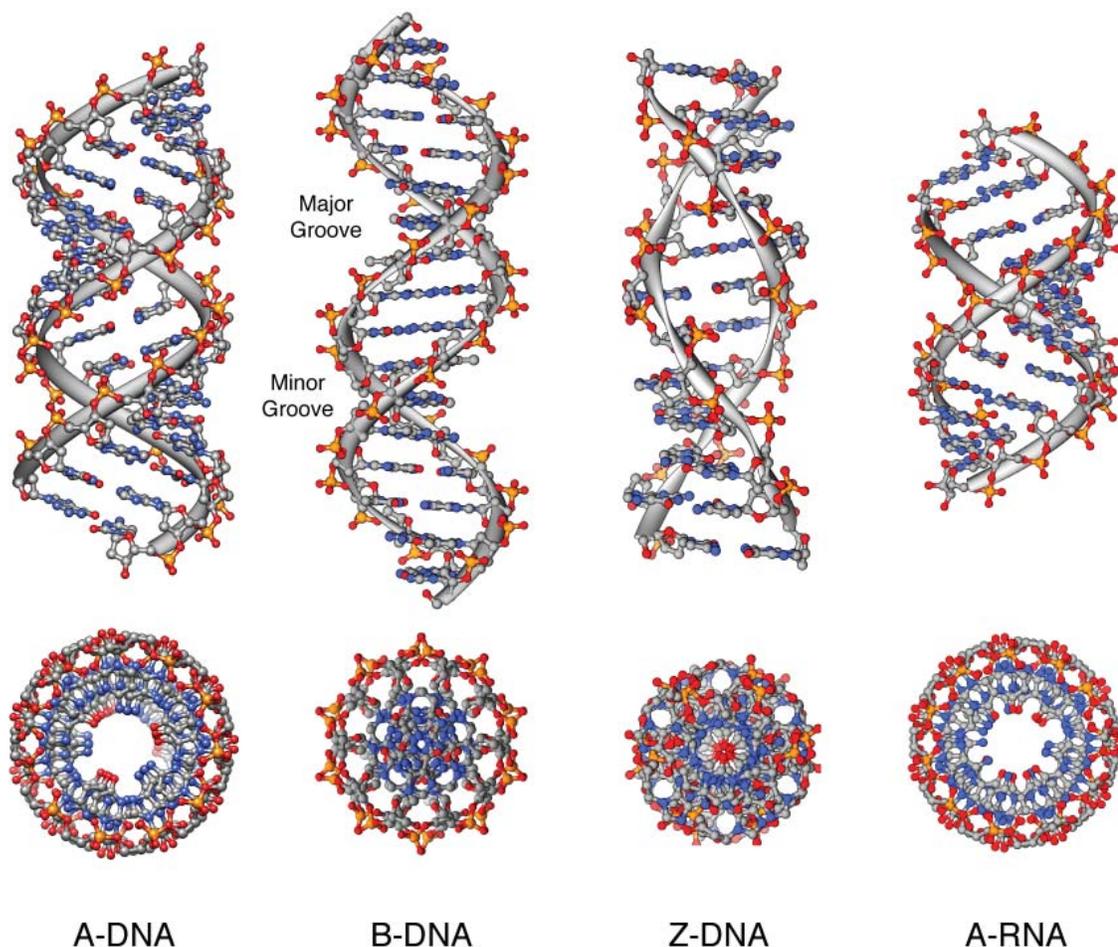
In biological systems, especially eukaryotic cells, DNA is assembled around octameric proteins called histones and compacted into macrohelical fibers forming the high-order structure of chromatin. This DNA-histone complex is called the nucleosome core particle (NCP) and represents the fundamental repeating unit of chromatin consisting of 147 base pairs of DNA forming two super helical turns around the histone octamer with 20-80 base pairs of linker DNA separating one NCP from the next. The Richmond group<sup>25-27</sup> at ETH Zurich has made seminal contributions to elucidate biologically relevant higher-order DNA structures such as the NCP<sup>25,26</sup> and the tetra-NCP<sup>27</sup> presented Figure 1.5. In addition, a theoretical model of four tetra-NCPs assembled into a super-helical chromatin fiber is presented in Figure 1.5. Chromatin architecture and accessibility in biological systems represents a higher-order level of regulation and a profoundly important problem for the field of DNA recognition.

### 1.3 Molecular Recognition of DNA

One of the largest projects in modern science, the human genome project,<sup>28-30</sup> is poised to deliver detailed information and make major impacts in biotechnology and medicine through the physical and functional characterization of the approximately 20,000 to 25,000 genes in the human



**Figure 1.3** DNA sugar phosphate backbone torsion angle map and sugar pucker conventions. a) Chemical structure of DNA sugar phosphate backbone with torsion angles next to the 3-dimensional structure of a nucleoside taken from the native high resolution B-DNA structure solved in Chapter 4 of this thesis. b) Pseudorotation phase angle ( $P$ ) diagram defining 5-membered ring sugar pucker modes. Equation describing the pseudorotation phase angle and maximum torsion angle for 5-membered rings. c) Chemical structure of the most common sugar pucker modes for B-DNA ( $C2'$ -endo) and A-DNA/A-RNA ( $C3'$ -endo).



**Figure 1.4** A comparison of double helical DNA polymorphs and A-form RNA.

genome. These genes are tightly regulated in higher organisms by transcription factor assemblies that function in a concerted cooperative and combinatorial fashion to modulate eukaryotic gene expression. The molecular recognition processes involved in nucleic acid-protein interactions are completely analogous to those of nucleic acid-drug interactions where initial complexation is often driven by the hydrophobic effect. Optimization of the same forces is also required, involving minimization of water exposed hydrophobic surfaces and maximization of van der Waals interactions in conjunction with the optimization of all buried hydrogen bond donor and acceptor pairings (solvent-assisted or counterion charge neutralization).<sup>31</sup> The recognition of the B-DNA interface by proteins and small molecules can occur at the major groove, minor groove, and phosphate backbone, or any combination, with interactions mediated through electrostatics, hydrogen bonding, and van der Waals interactions along with base pair stacking for the case of intercalators. The DNA base pair edges in the major groove and minor groove provide an array

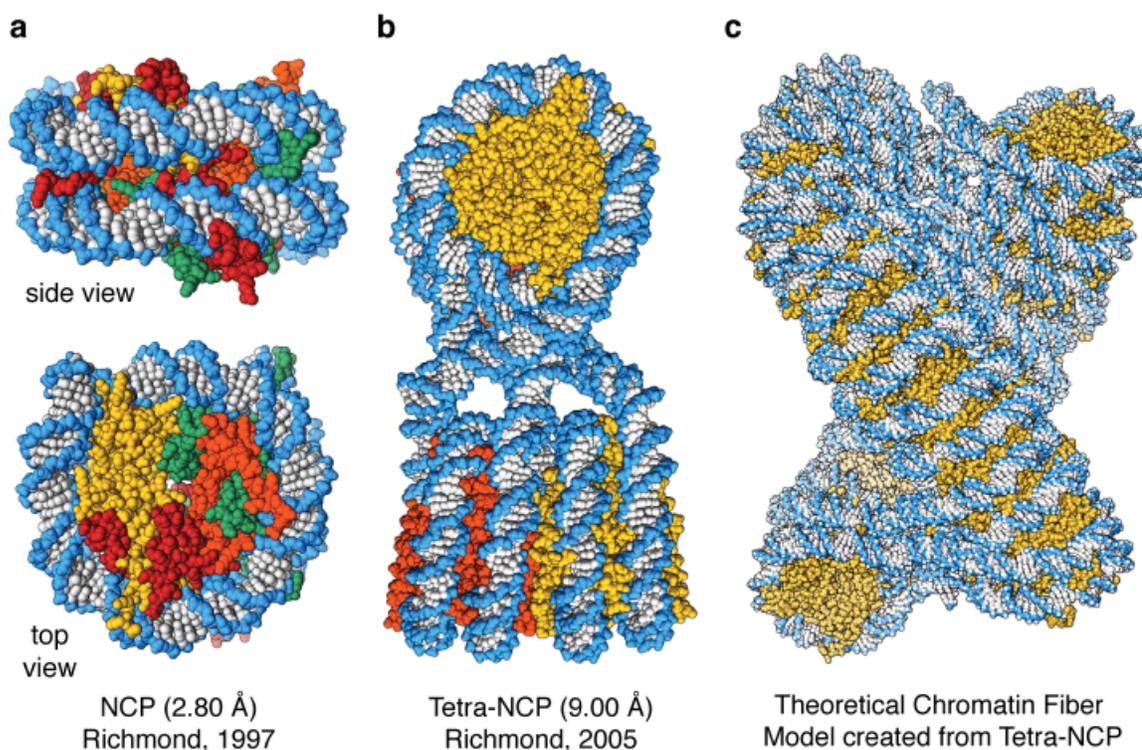
**Table 1.1** Typical nucleic acid structural parameters.

	A-DNA	B-DNA	Z-DNA	A-RNA
Helix Sense	Right	Right	Left	Right
Screw symmetry	11-fold	10-fold	6-fold	12-fold
bp/repeating unit	1	1	2	1
bp/turn	11	10	12	11-12
Helix twist	32.7°	36.0°	-10 <sup>a</sup> , -50 <sup>b</sup>	33.1°
Rise (Å)	2.8	3.4	-3.9 <sup>a</sup> , -3.5 <sup>b</sup>	2.8
Helix pitch (Å)	28	34	44	36
Base pair tilt (Å)	20	-6	7	16.7
Diameter of helix (Å)	23	20	18	21
Rotation per bp	33	36	-30	32.7
Glycosidic bond				
dA, dT, dC	anti	anti	anti	anti
dG	anti	anti	syn	anti
Sugar pucker				
dA, dT, dC	C3' endo	C2' endo	C2' endo	C3' endo
dG	C3' endo	C2' endo	C3' endo	C3' endo
Phosphate-phosphate (Å)				
dA, dT, dC	5.9	7.0	7.0	~5.7
dG	7.0	7.0	5.9	~5.7
Major groove				
width (Å)	2.7	11.7	Convex	11.1
depth (Å)	13.5	8.5		
Minor groove				
width (Å)	11.0	5.7	4	11.1
depth (Å)	2.8	7.5	9	shallow

<sup>a</sup> CpG step. <sup>b</sup> GpC step.

of functionality for hydrogen bonding, hydrophobic interaction, and steric complementarity with proteins and small molecule binders.<sup>5,8,16,17,31</sup> The molecule electrostatic potential surfaces for the minor and major groove base pair edges are shown in Figure 1.6.<sup>32</sup> In addition, primary driving forces such as the hydrophobic effect and shape complementarity are common to both proteins and small molecules.

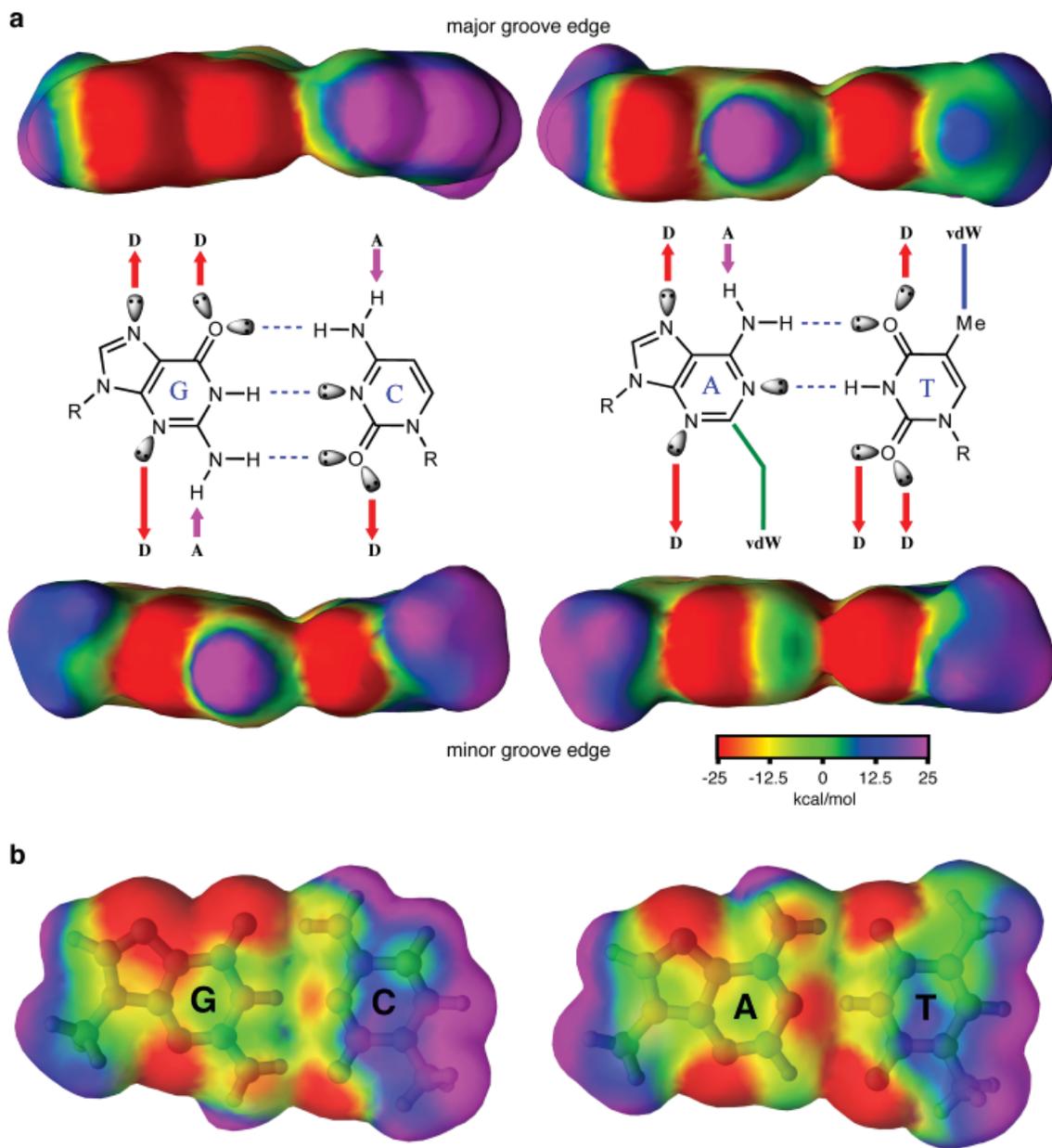
The regulation of gene transcription is controlled by the sequence specific cooperative assembly of transcription factors, which form regulatory switches and networks in the cell providing stringent control over biochemical processes.<sup>33</sup> The minor groove and major groove of



**Figure 1.5** Nucleosome core particle structures. a) Crystal structure of the nucleosome core particle at 2.80Å determined by Richmond and coworkers. b) X-ray structure of the tetra-NCP determined at 9Å resolution. c) Theoretical model of a chromatin fiber constructed from four tetra-NCPs by rotation and translation about the central axis.

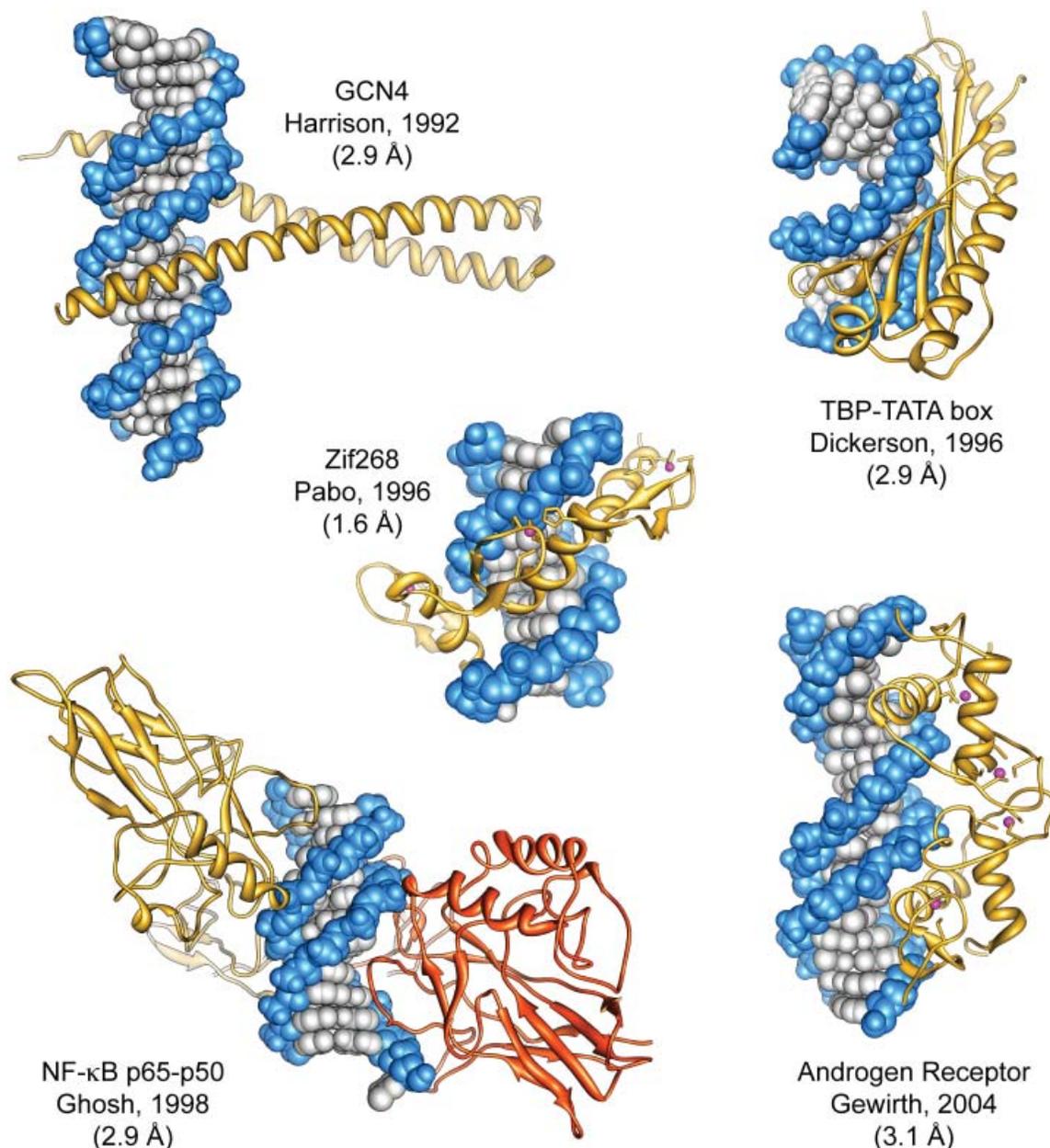
DNA provide distinct surfaces for the interaction of transcription factors through specific and nonspecific interactions such as (hydrogen bonding, electrostatics, van der Waals, etc.). Several DNA-binding transcription factors are presented in Figure 1.7 to highlight the diverse architectures used for recognizing DNA, ranging from homodimeric coiled coils interacting with the major groove to monomeric beta-sheet containing proteins interacting with the minor groove. In addition to homodimeric motifs, heterodimeric motifs are utilized along with metal ion coordinated assemblies (i.e. Nf- $\kappa$ B p65-p50 and androgen receptor).<sup>33</sup> Transcriptional co-activating proteins serve to integrate information from transcription factor assemblies and modulate gene expression through communication with RNA polymerase II leading to the transcription of protein-coding regions in the eukaryotic genome.<sup>33,34</sup>

Transcription factors (TF) can communicate indirectly through allosteric modulation of DNA resulting in cooperative assembly with very little direct protein-protein interaction. Transcription factor binding can cause DNA-sequence dependent structural perturbations which modulate the binding of the next TF. TF's can also interact directly through protein-protein interactions to increase



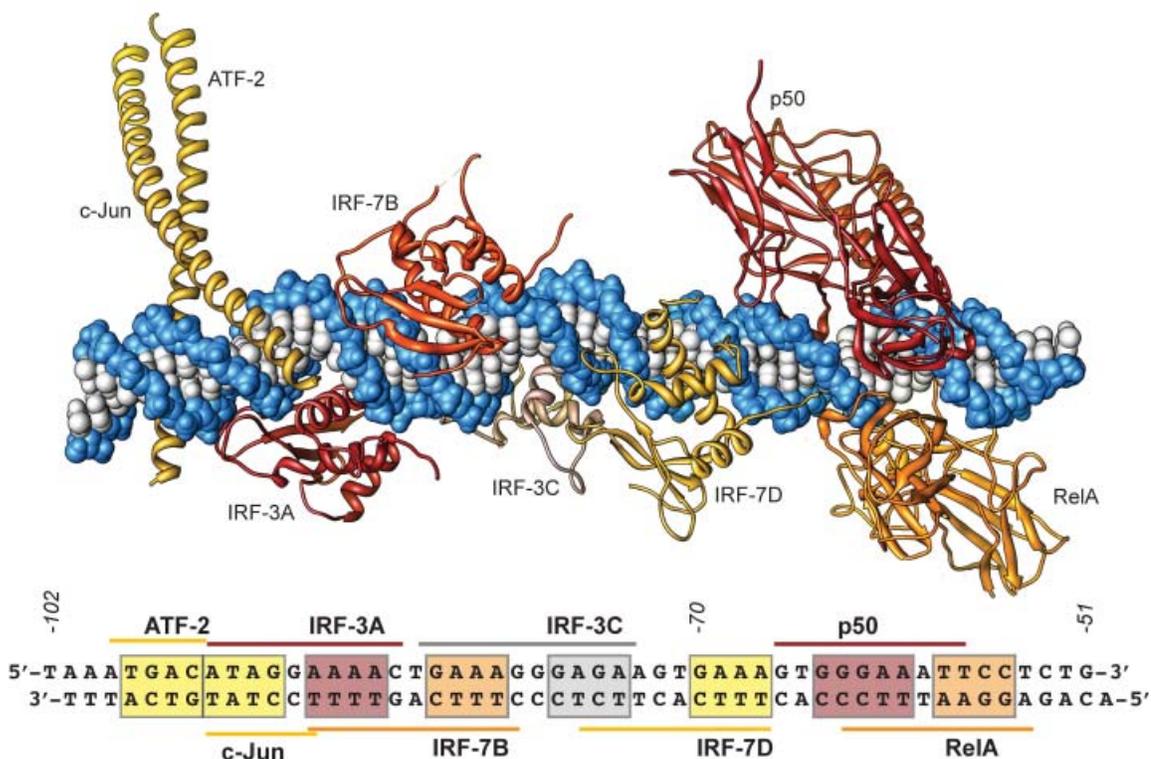
**Figure 1.6** Anatomy of the DNA base pair edges and their molecular recognition properties. a) Electrostatic potential maps of the base pair edges presented to the major (top) and minor (bottom) grooves of DNA. Hydrogen bond donors are designated with red arrows and the letter D. Hydrogen bond acceptors are designated with purple arrows and the letter A and functionality for the potential for van der Waals interactions is designated with the appropriate colored line and vdW. b) Top view of the Watson-Crick base pair molecular electrostatic potential surfaces. Electrostatic potential maps from native DNA crystal structure solved in Chapter 4 of this Thesis.

cooperativity. The  $\beta$ -enhancesome (Figure 1.8) is one such example of a cooperative assembly with cooperativity most likely arising at the DNA and coactivator levels. A conserved stretch of 55 bp's



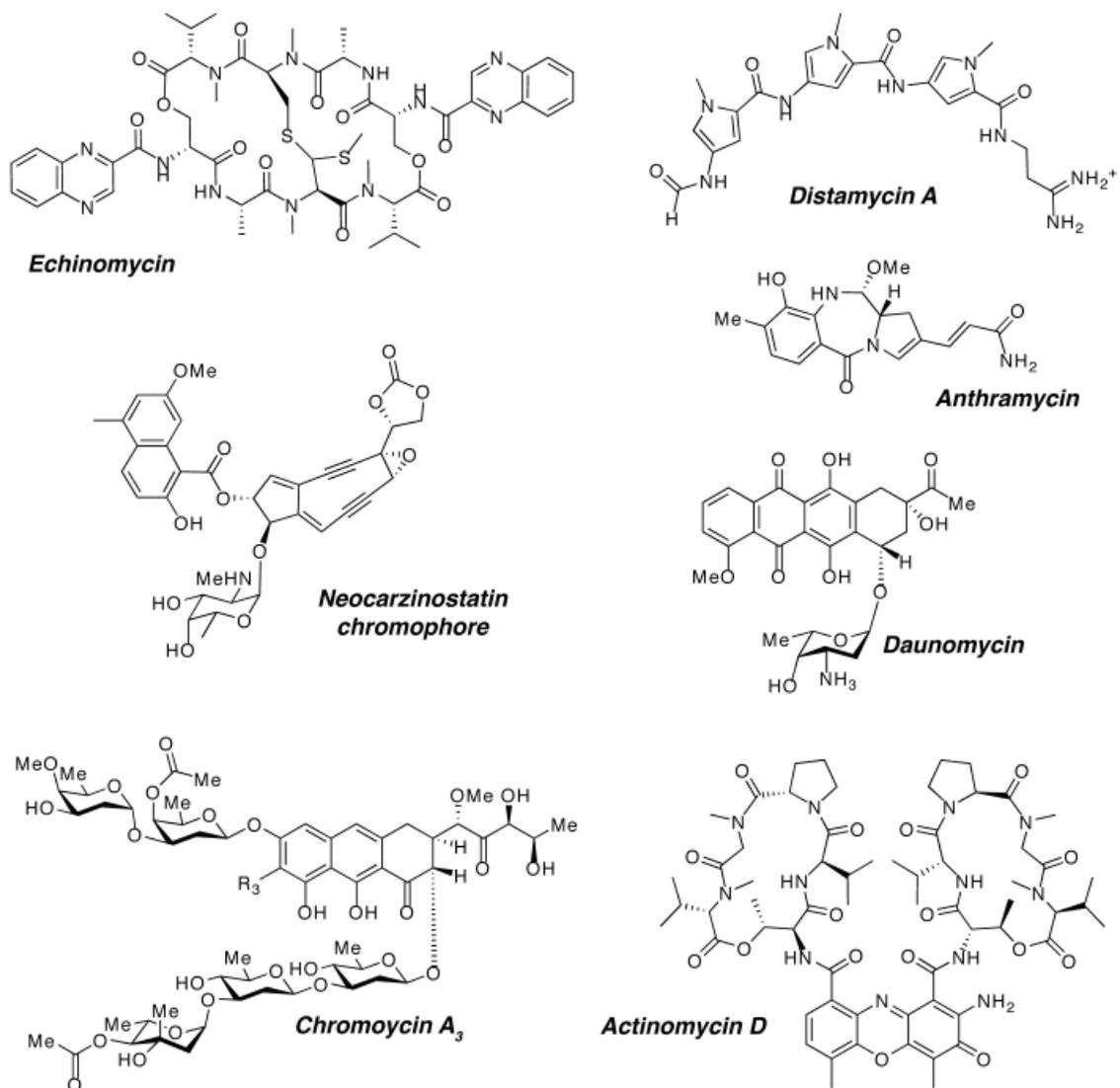
**Figure 1.7** X-ray structures of DNA binding transcription factors [GCN4 (Harrison, 1993), Nf- $\kappa$ B p65-p50 (Ghosh, 1998), TBP-TATA box (Dickerson, 1996), Zif268 (Pabo, 1996), Androgen receptor (Gewirth, 2004)].

(160 Å long) in a nucleosome free region of the IFN- $\beta$  promoter serves as a regulatory element for the cooperative assembly of 8 proteins into a continuous surface, burying 72% of the DNA solvent accessible area with very little protein-protein interaction.<sup>35</sup> Transcriptional co-activating proteins serve to integrate information from the assembly to modulate gene expression through communication with RNA polymerase II leading to transcription.<sup>33,34</sup>



**Figure 1.8** Atomic model of the cooperative assembly of interferon- $\beta$  enhancer showing 4-6 base pair transcription factor binding sites along the highly conserved composite DNA interface of 55 base pairs spanning approximately 160 Å in length.<sup>35</sup>

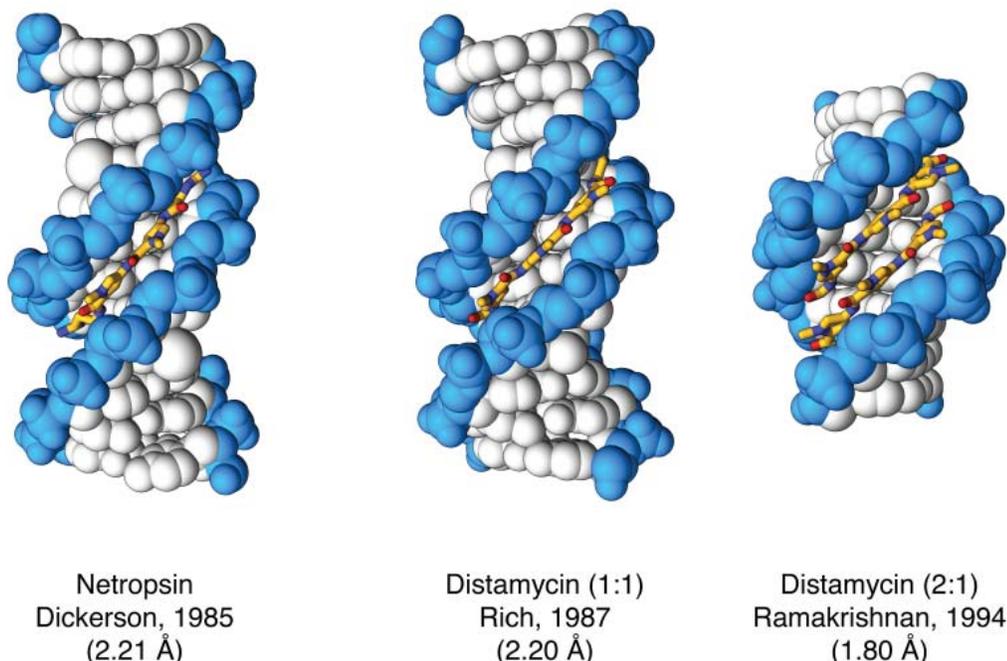
A diverse range of natural products and secondary metabolites have been shown to bind DNA with interaction modes consisting primarily of either intercalation or groove binding.<sup>5,7,8</sup> In addition, some ligands rely on a combination of intercalation and groove binding that can also be augmented by covalent modifying chemical domains, as in the case of anthramycin and neocarzinostatin. A collection of diverse DNA binding natural products are shown in Figure 1.9 with echinomycin and daunomycin representing intercalators and anthramycin and distamycin A representing minor groove binders. The natural product distamycin (Figure 1.9) binds to A,T tracks in the minor groove of DNA, four to five base pairs in size, in both a 2:1 and a 1:1 ligand:DNA stoichiometry.<sup>12-14</sup> The affinity and specificity of distamycin is controlled by a superposition of shape complementarity, hydrophobic effects, and specific hydrogen bonding to the minor groove of B-form DNA. Due to its modular design of repeating pyrrole amino acids and amenability to rational modification, distamycin has served as the inspiration for the design of several classes sequence specific DNA minor-groove binders, with the ultimate goal of designing highly specific targeted gene regulation agents.



**Figure 1.9** DNA-binding natural products.

#### 1.4 DNA Recognition by Minor-Groove Binders

Prior to the first structure of a molecule bound to DNA, specific recognition of B-form DNA was predicted to occur in major groove rather than minor groove.<sup>5</sup> An observation that was based on the fact that the hydrogen bond acceptors at N3 of adenine and O2 of thymine A/T base pairs are similarly placed and lack any prominent distinguishing features.<sup>36</sup> With the combination of biophysical and structural data from NMR and X-ray studies, it was verified that the minor groove of B-form DNA was a legitimate target for specific recognition.<sup>12-15</sup> (For crystal structures of netropsin and distamycin A, see Figure 1.10.) Building upon inspiration from the natural products, netropsin



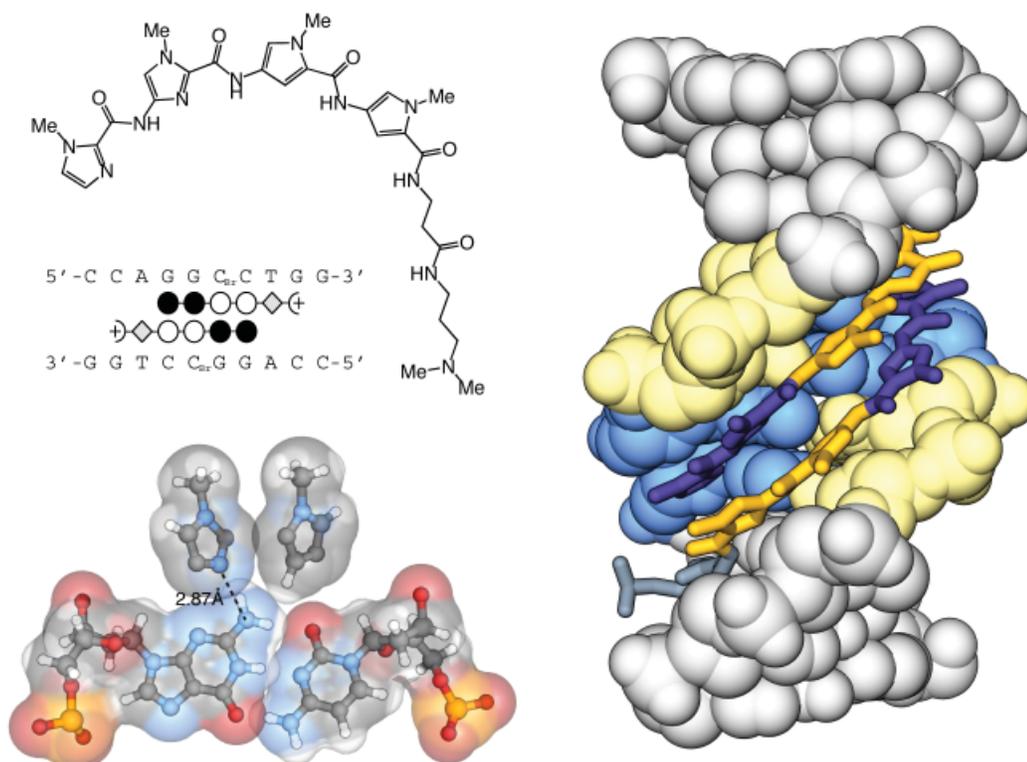
**Figure 1.10** X-ray crystal structures illustrating DNA recognition by the natural products netropsin (left) and distamycin (middle, right). The dicationic natural product netropsin binds preferentially to narrow AT tracts in the DNA minor groove as a monomer (PDB: 6BNA). The monocationic natural product distamycin also binds AT tracts of the DNA minor groove in a 1:1 (PDB: 2DND) and 2:1 (PDB: 378D) ligand:DNA stoichiometry.

and distamycin, minor groove binders have progressed to a modular molecular recognition platform with high affinity and specificity for many different sequences of DNA.<sup>5,7,8,16,17</sup>

Over the past two decades, the development of minor groove DNA binders has evolved from the initial discovery of the natural product distamycin to a new class of programmable heterocyclic oligomers demonstrating high affinity and sequence specificity.<sup>16,17</sup> In addition to the incorporation of alternative heterocycles such as imidazole that have enabled specificity for guanine recognition using the Im-Py pair, much research has gone into linking the two heterocyclic strands in a dimeric motif.<sup>37-39</sup> Covalent linkage of the two anti-parallel heterocyclic strands by a gamma amino butyric acid (GABA) linkage results in increases in affinity of 100–3600 fold relative to the unlinked homodimeric motif.<sup>40,41</sup> The incorporation of the turn linkage in the form of a GABA or substituted GABA turn represented a major technological advance allowing for the first time the incorporation of unsymmetrical ring pairs for the targeting of non-palindromic DNA sequences.<sup>37</sup> In addition, covalent linkage of the two strands has led to sub-nanomolar increases in affinity competing with and often rivaling that of endogenous DNA binding proteins.<sup>16,17,37-39</sup> This high affinity modular dimeric motif has allowed for the regulation of gene expression by direct interaction with the DNA-

protein interface.<sup>16,17</sup>

The four Watson-Crick base pairs can be differentiated by their molecular shape, electrostatic potential, and positions of hydrogen bond donors and acceptors in the DNA minor groove floor. The minor groove edge of a G•C base pair contains a steric hydrogen bond donating “bump” in the form of the exocyclic amine of guanine. The steric properties of the exocyclic amine of guanine form the basis for the A•T selectivity observed for netropsin and dystamycin binding due to steric interaction with the edge of the pyrrole ring. It was discovered in a key study in the early 1990s that imidazole in place of pyrrole in a three ring polyamide analogous to dystamycin could bind the 5' base pair sequence 5'-WGWCW-3' (where W=A or T) resulting in a 2:1 polyamide-DNA complex where the imidazole ring is stacked against a pyrrole ring allowing differentiation of G•C base pairs

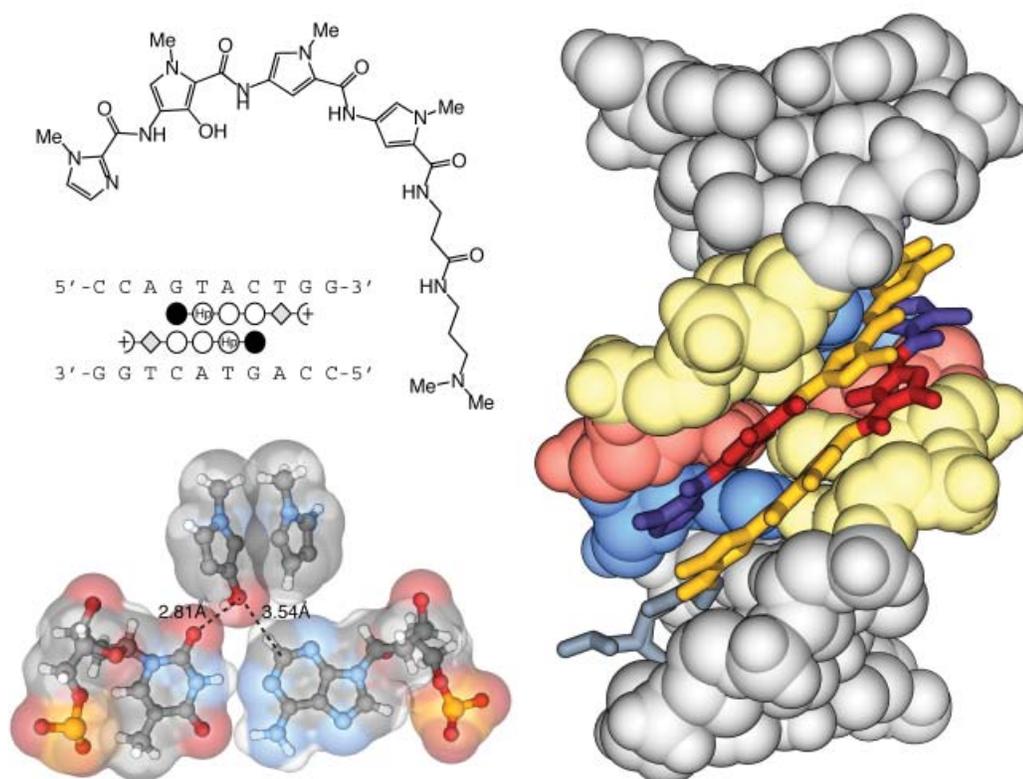


**Figure 1.11** Crystal structure (PDB: 365D) of the 2:1 binding single strand Py/Im polyamide targeted to the sequence 5'-CCAGGCCTGG-3' (2.00 Å resolution). Overall complex is shown on the right and the space filling model showing the basis for GC recognition is at the bottom left where the imidazole lone pair forms a hydrogen bond with the exocyclic N-H of guanine.

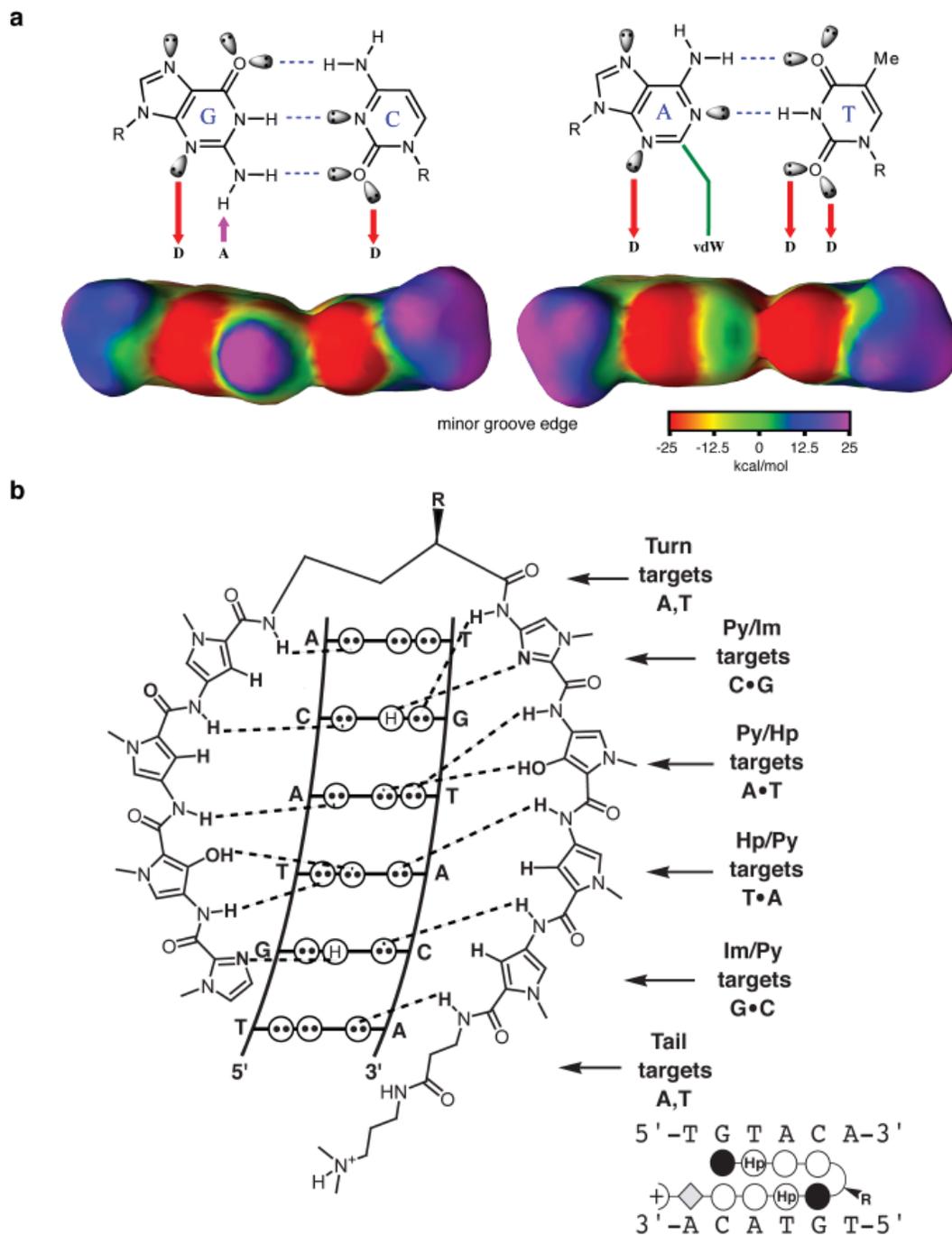
from C•G, A•T, or T•A.<sup>42</sup> The Im/Py pair has been used extensively in unlinked polyamides and in turn linked polyamides culminating in the recent publication of a polyamide library, that represents

the solutions for targeting various 6 base pair sequences with high affinity and specificity.<sup>16,17,43</sup> Thermodynamic studies have revealed that the Im/Py pair sequence selectivity is primarily driven by favorable enthalpic factors<sup>44,45</sup> and X-ray crystallographic studies in collaboration with the Rees group provided structural insight based on a specific hydrogen bond between the imidazole lone pair and the exocyclic amine of guanine in unlinked 2:1 homodimeric polyamides (Figure 1.11).<sup>46</sup>

Discrimination of T•A from A•T base pairs represents a much greater challenge due to the ability of thymine and adenine to both accept a hydrogen bond and the lack of unsymmetrical steric features as in the G•C case.<sup>16,17</sup> Despite this challenge, a small asymmetric cleft between the C2 of adenine and the O2 of thymine has been exploited for specific targeting by the *N*-methyl-3-hydroxypyrrrole/*N*-methylpyrrole (Hp/Py) pair, however affinities of these molecules are slightly lower than their Py/Py containing counterparts.<sup>16,17</sup> In another seminal structural study with the Rees group on Hp containing 2:1 binders, it was revealed that a combination of shape selective recognition of the asymmetric cleft along with a specific hydrogen bond between the Hp hydroxyl and the thymine O2 was responsible for the A•T specificity (Figure 1.12).<sup>47,48</sup> The combination of



**Figure 1.12** Crystal structure (PDB: 407D) of the 2:1 binding single strand ImHpPyPy-β-Dp polyamide targeted to the sequence 5'-CCAGTACTGG-3' (2.20 Å resolution). Overall complex is shown on the right and the space filling model of the basis for AT recognition is shown at the bottom left where the hydroxypyrrrole-OH is within hydrogen bonding distance of the carbonyl lone pair on T.

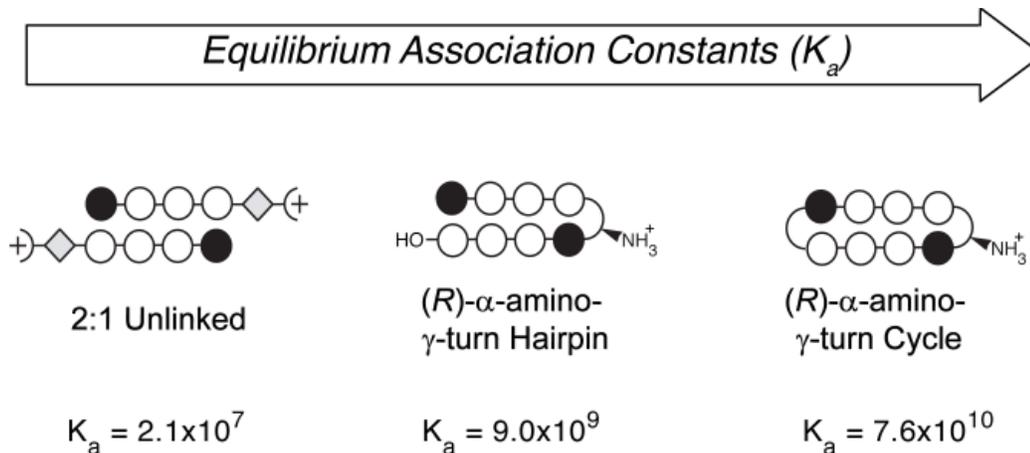


**Figure 1.13** Molecular recognition of the DNA minor groove and polyamide pairing rules. a) Molecular recognition of the DNA minor groove edges with molecular electrostatic potential surfaces showing the array of hydrogen bond donors (D), hydrogen bond acceptors (A), and hydrophobic functionality (vdW). b) Polyamide pairing rules.

Py, Im, and Hp combined as unsymmetrical pairs in opposite strands of a unlinked homodimeric or turn linked polyamide can be used to specifically recognize the four Watson-Crick base pairs

(Figure 1.13).<sup>16,17</sup> These interactions can be described as a set of guidelines or pairing rules for the design of sequence specific B-form DNA targeted polyamides where Im/Py specifies G•C and Hp/Py specifies A•T. Some limitations do exist for certain sequences such as homopurine tracts, certain G-rich sequences, and sequences beyond 6 base pairs due to the sequence-dependent DNA microstructure and overcurvature of longer polyamides, however unique solutions to some of these problems have been developed (i.e. incorporation of a flexible  $\beta$ -Ala residue in 1:1 and hairpin polyamide motifs).<sup>16,17</sup>

In addition to the 2:1 and hairpin polyamide architectures many other strand linkage strategies have been explored such as linking through the *N*-methyl groups on the central heterocycles (H-pin motif)<sup>49-50</sup> or in the terminal heterocycles (U-pin motif).<sup>51</sup> However, one of the highest affinity and in some cases most specific polyamide architectures has been the covalent linking of the C- to N-termini at both ends of the polyamide into a macrocycle, eliminating all possibility of extended binding modes.<sup>41,52-55</sup> Macrocytic  $\gamma$ -turn linked polyamides were first explored as 6 ring systems targeting a 5 base pair sequence of DNA in 1995 and were shown to have significantly higher affinity, however their specificity versus mismatch DNA was only 3-fold compared to 40-fold for their hairpin counterparts.<sup>52</sup> Mainly due to limitations in synthetic methodology and initial

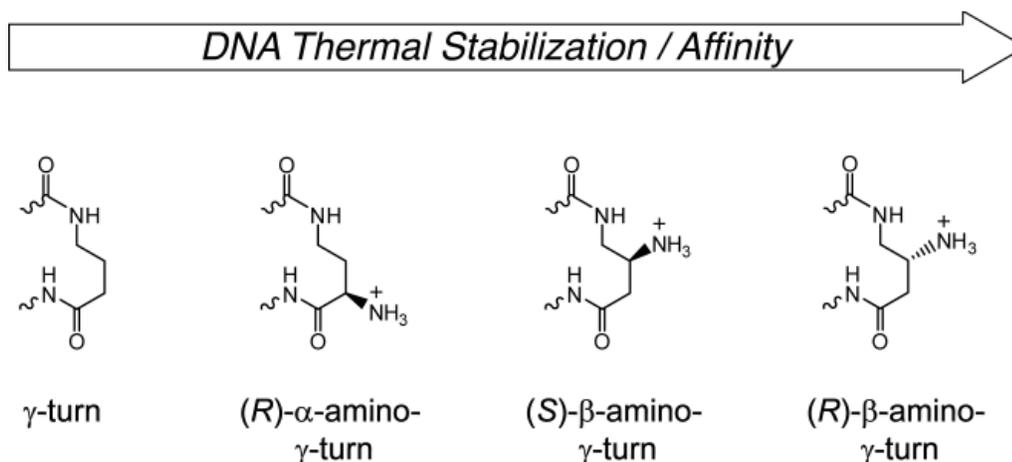


**Figure 1.14** Consequence of covalent attachment of two polyamide strands by incorporation of GABA-based turns. (For a structural key to the ball-and-stick nomenclature see the Nomenclature and Symbology section at the beginning of this thesis or Figure 1.13).

discouraging thermodynamic results the cyclic polyamide motif was not investigated further until 1999.<sup>41</sup> After improvements in solid-phase synthetic methodology, nanomole to micromole quantities of polyamides could be readily synthesized although cyclic polyamides still remained

challenging.<sup>53-55</sup> Using solid-phase methods cyclic polyamides were reinvestigated with two major architectural changes.<sup>41,53-55</sup> The first being the use of an 8 rings system as oppose to 6 in the original studies and the second major change was moving the charge from the pyrrole *N*-methyl group to the alpha position of the  $\gamma$ -turn in the form of (*R*)-2,4-diaminobutyric acid, that had been discovered to increase the affinity, sequence specificity, and orientational preference of hairpin polyamides. This second generation cyclic 8 ring polyamide motif was found to have greatly improved specificity and affinity over its hairpin and unlinked counterparts targeting the sequence 5'-AGTACT-3'.<sup>41</sup> The results of covalent attachment of the two polyamide strands can be seen in Figure 1.14. In a second study, multiple Hp/Py pairs were introduced into the 8-ring cyclic polyamide motif, which resulted in increased affinity and specificity relative to hairpin polyamides targeted to the sequences 5'-TGAACT-3' and 5'-TGATCT-3'.<sup>53</sup> Despite these advances the cyclic polyamide motif has received little attention relative to its hairpin counterpart mainly due to synthetic limitations.

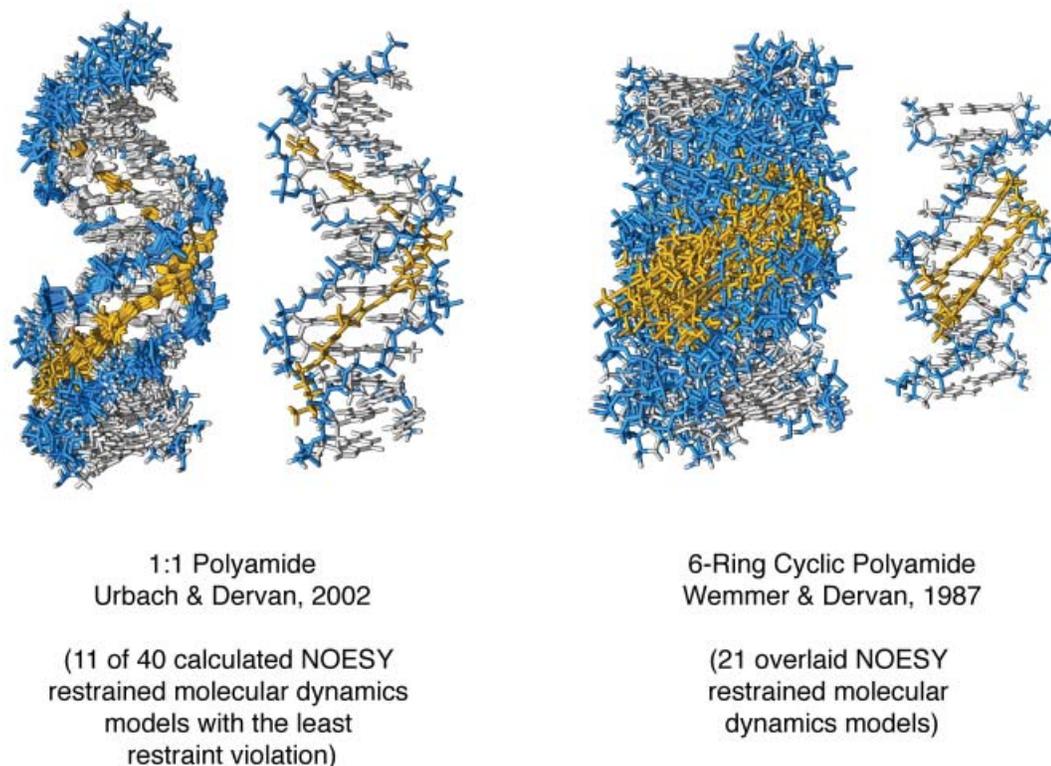
The chiral (*R*)-2,4-diaminobutyric acid turn ( $\alpha$ -turn) was a major advance in polyamide design not only for the cyclic polyamides but primarily for the hairpin motif.<sup>38</sup> The addition of an amino substituent to the alpha position of the  $\gamma$ -turn helps to disfavor extended 1:1 binding



**Figure 1.15** Consequence of covalent attachment of two polyamide strands by incorporation of GABA-based turns.

modes and reverse binding due to a steric clash with the minor groove floor. In addition, the chiral amino turn helps to increase the overall affinity of polyamides while maintaining specificity and improving water solubility.<sup>38</sup> The chiral  $\alpha$ -turn was proposed to increase affinity through electrostatic interactions between the protonated cationic amine group and the anionic DNA backbone however this interaction has not been born out in structural studies (see Chapter 5 and 6 of this thesis).<sup>17</sup> In

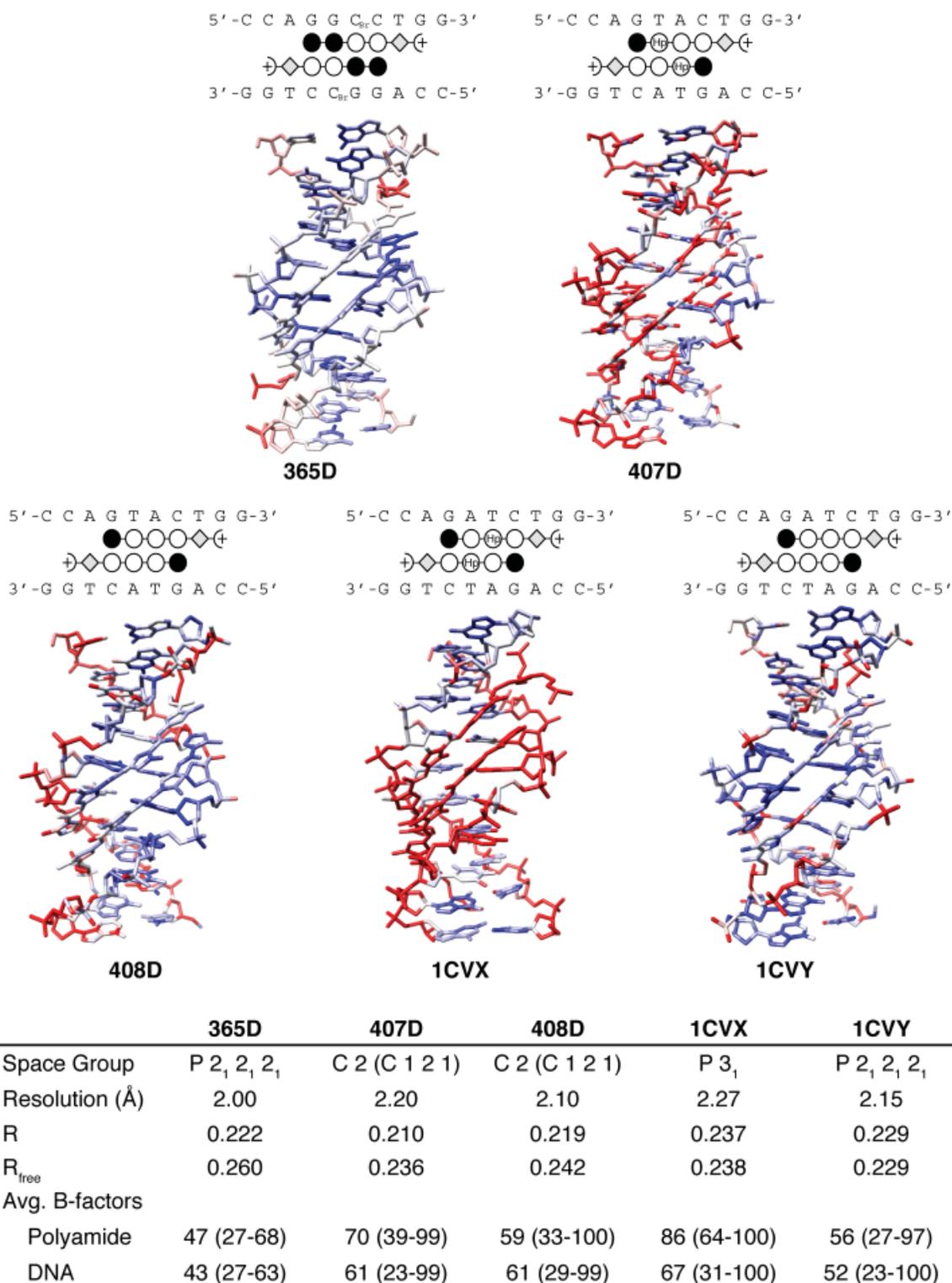
addition to substitution of the  $\gamma$ -turn at the alpha position, many other variations have been studied with shorter turns, longer turns, and conformationally constrained variations.<sup>16,17,56</sup> The most recent and one of the most successful advances in polyamide turn technology is the  $\beta$ -amino  $\gamma$ -linked



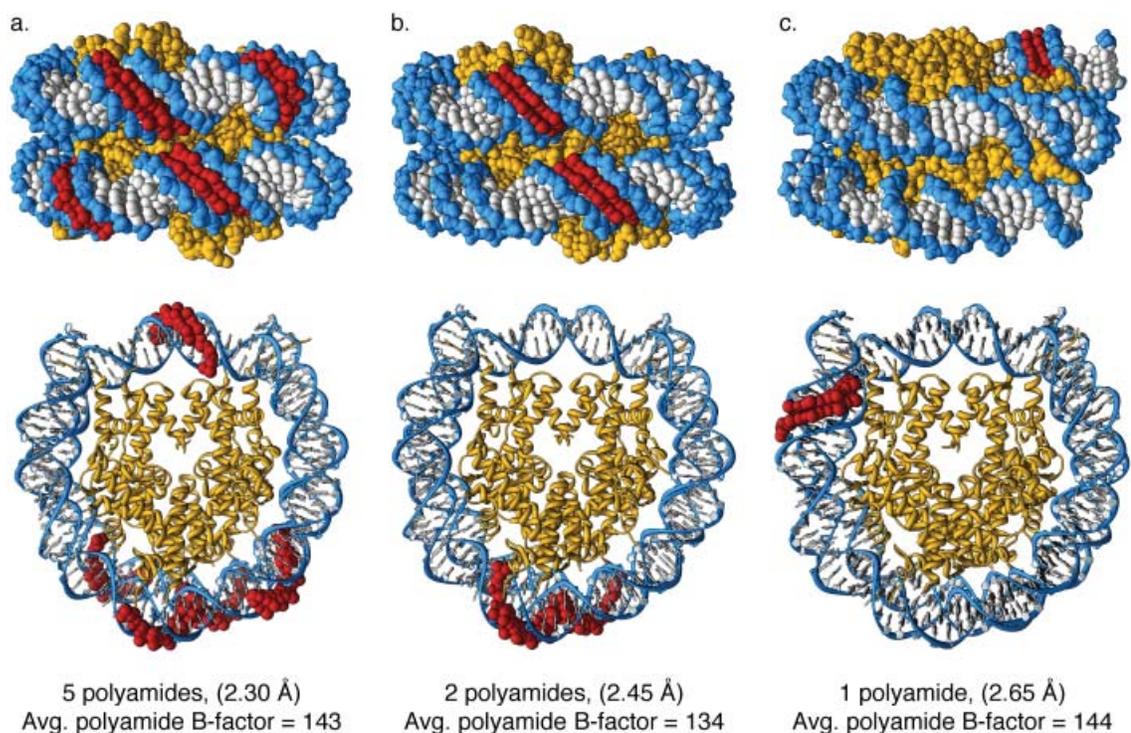
**Figure 1.16** NMR NOESY-restrained molecular dynamics models of 1:1 and 6-ring cyclic polyamide DNA complexes.

turn.<sup>39</sup> Recent studies have demonstrated that this turn can provide substantial increases in affinity for certain polyamide sequences, however the effect is less pronounced as the imidazole content of the polyamide is increased (Figure 1.15).<sup>39</sup>

NMR structural studies using NOESY-restrained molecular dynamics models have also provided insight into 1:1 and 6-ring cyclic polyamides complexed with DNA.<sup>55,57</sup> Figure 1.16 shows an NMR model of a 1:1 polyamide with 11 out of 40 of the best calculated models overlaid.<sup>57</sup> This shows very small coordinate deviation towards the center of the DNA helix and bound polyamide with increasing conformational mobility at the ends. The 6-ring cyclic polyamide model represented in Figure 1.16 shows an overlay of 21 of the best calculated models.<sup>55</sup> This structure shows significant conformational mobility in the 6-ring cyclic complex with a highly flexible DNA sugar-phosphate backbone.

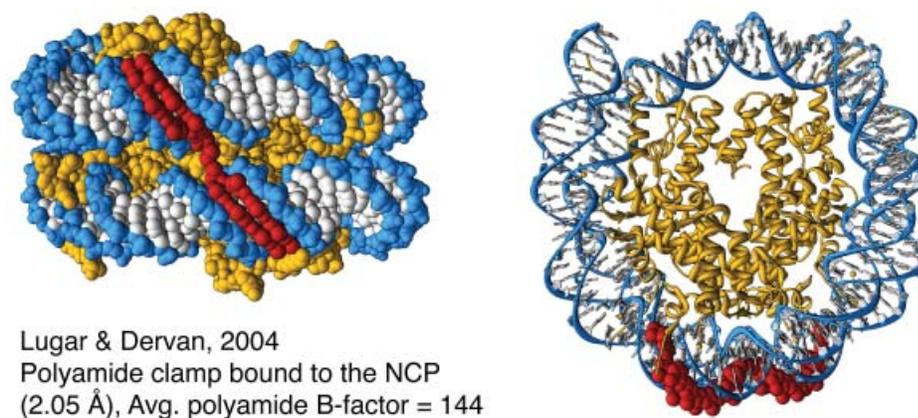


**Figure 1.17** Polyamide 2:1 DNA crystal structures colored by B-factor with red representing the largest B-factors and blue representing the smallest.



**Figure 1.18** X-ray crystal structures of polyamide-NCP complexes. a) Five polyamides in complex with the NCP at 2.30 Å resolution. b) Two polyamides in complex with the NCP at 2.45 Å resolution. c) One polyamides in complex with the NCP at 2.65 Å resolution.

X-ray crystallographic studies resulting from collaborations between the Rees and Dervan groups have provided valuable insight into the polyamide-DNA molecular recognition process by elucidating the structure of five 2:1 polyamide-DNA complexes at a resolution ranging from 2.00 to 2.27 Å and a summary comparing specific structural parameters is shown in Figure 1.17.<sup>46-48</sup> These crystallographic studies revealed that the DNA rise per base pair matches the polyamide rise



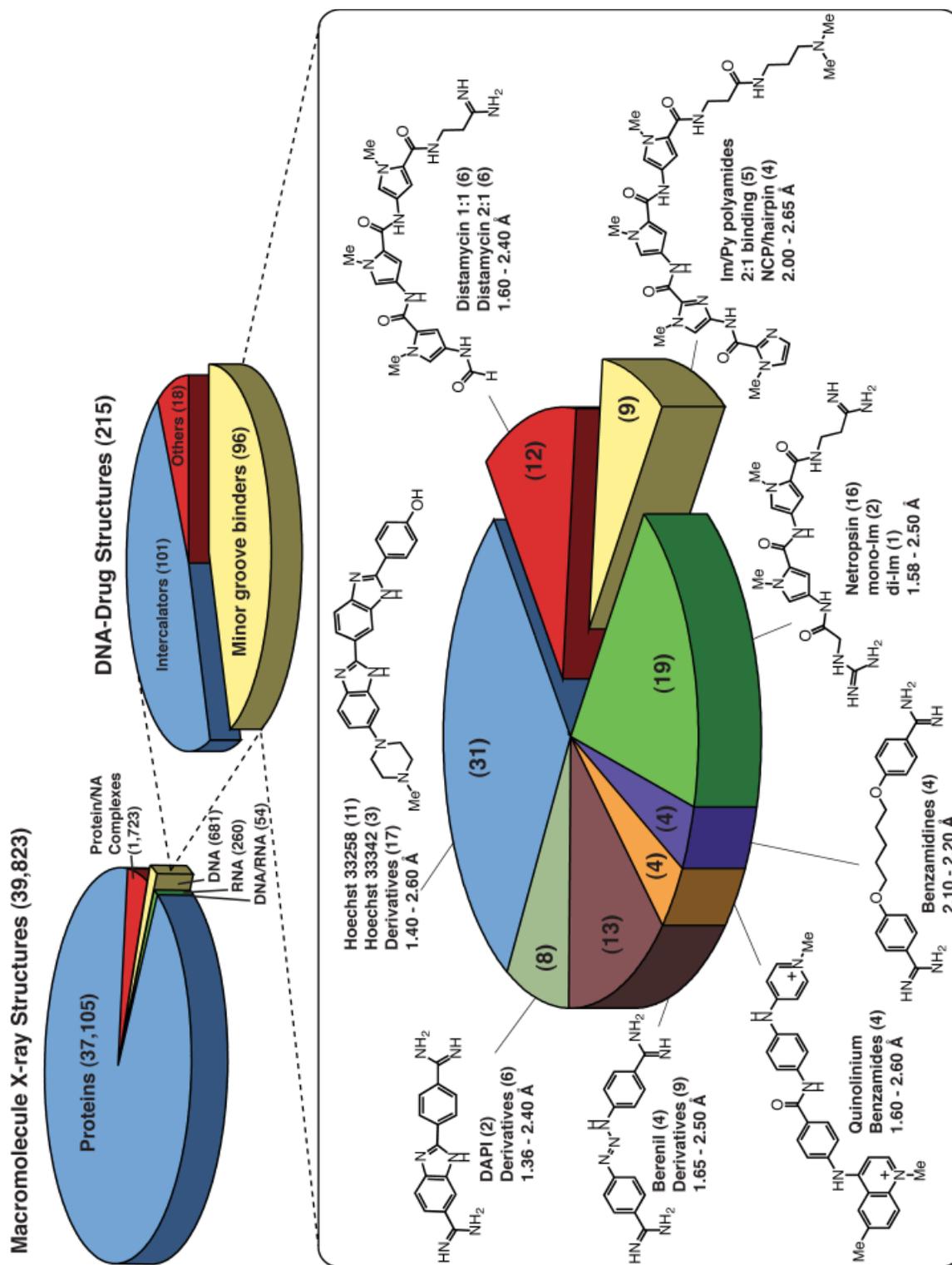
**Figure 1.19** Turn-linked polyamide clamp bound to the NCP with the linker traversing the nucleosome super-groove.

per residue, however the polyamide structure is over-curved with respect to the DNA minor groove and shape complementarity is lost beyond a sequence of 5 contiguous base pairs.<sup>46</sup> In addition, these crystallographic studies elucidated the basis for GC recognition by the Im/Py pair<sup>46</sup> and TA recognition by the Hp/Py pair<sup>47,48</sup> providing fundamental insight into polyamide binding. A variety of space groups were observed including monoclinic, orthorhombic, and trigonal with resolutions ranging from 2.00 to 2.27 Å. Average R-factors were in the mid-20s and polyamide B-factors averaged 47 to 86 Å<sup>2</sup> whereas DNA B-factors averaged 43 to 67 Å<sup>2</sup> for all structures presented in Figure 1.17. The 2:1 binding polyamide crystal structures also frequently exhibited disorder in the polyamide tail region and was usually modeled in alternate conformations reflecting the dynamic nature of the β-alanine linked dimethylamino propylamine terminus.

DNA binding polyamides are also able to bind sequence specifically to DNA on the nucleosome core particle.<sup>58</sup> Hairpin polyamide-NCP crystal structures have been solved at resolutions ranging from 2.05 to 2.65 Å providing structural proof that polyamides can bind biologically relevant higher-order DNA structure however a combination of resolution limits and high B-factors for the polyamide prevented a detailed picture beyond confirmation of the polyamide binding location (Figures 1.18 and 1.19).<sup>58,59</sup> The current state of macromolecular crystallography, with regard to minor groove binding DNA-drug structures, was assessed prior to beginning the structural work presented in Chapter 5 and 6 of this thesis and is presented in Figure 1.20. This survey demonstrates the lack of high resolution structures of DNA minor groove binders and the notable absence of linked dimeric minor groove binder structures. This survey underscores the pressing need for atomic resolution X-ray crystal structures of DNA minor groove binders to truly understand the molecular basis of recognition.

## 1.5 Scope of this work

The work presented in this thesis is focused on the molecular recognition of DNA by minor groove binding polyamides. In Chapter 2 of this thesis, a solution-phase synthesis of pyrrole-imidazole polyamides is presented with optimized protocols utilizing little to no chromatography. Chapter 3 builds on synthetic methodology in Chapter 2 allowing the efficient synthesis of cyclic polyamides targeted to the androgen response element. This chapter demonstrates that cyclic polyamides can be synthesized in an efficient manner, are biologically active and cell permeable in cell culture experiments, and rival the binding affinity of most other polyamide architectures. Chapter 4 details an oligomerization route to macrocyclic polyamides and reports on the DNA



**Figure 1.20** Current state of macromolecular crystallography: A DNA-drug perspective. Data compiled from the PDB on 11/04/2007.<sup>60</sup> (The number of structures solved is designated in parenthesis.)

binding ability of higher order macrocycles. The structural elucidation of an  $\alpha$ -amino-turn-linked cyclic polyamide is presented in Chapter 5 at 1.18 Å resolution providing insight into the detailed molecular recognition processes. Chapter 6 details the structural elucidation of a  $\beta$ -amino-turn-linked cyclic polyamide highlighting the conformational differences compared to the  $\alpha$ -amino-turns and providing a structural basis for the inability of polyamides to bind dsRNA. In Chapter 7, a new class of programmable oligomers targeting the DNA sequence 5'-WGGGGW-3' were shown to inhibit DNA binding of the Nf- $\kappa$ B transcription factor by EMSA gel shift. Compounds discovered in Chapter 7 were found to possess unique fluorescent properties with the ability to modulate their fluorescence by binding their targeted dsDNA site and this work is presented in Chapter 8. Chapter 9 describes an ongoing effort in the templated-assembly of polyamides using higher-order DNA structure (NCP). Additionally, this chapter describes the development of a new profluorescent class of heterocycle, that has the potential to be used as a chemical reporter for templated ligation events. Appendix A through F detail results from efforts not covered in the main thesis and a continuation of studies from Chapters 3 and 8.

## 1.6 Notes and Reference

1. Lehn, J. M. *Supramolecular Chemistry: Concepts and Perspectives* Wiley-VCH: Weinheim, Germany, **1995**.
2. Cram, D. J. The design of molecular hosts, guests, and their complexes *Science* **1988**, *240*, 760-767.
3. Cram, D. J. Molecular container compounds. *Nature* **1992**, *356*, 29-36.
4. Whitesides, G. M., Snyder, P. W., Moustakas, D. T., Mirica, K. A. Designing ligands to bind tightly to proteins. In *Physical Biology: From Atoms to Medicine* Zewail, A., Eds.; Imperial College Press: London, **2008**, 189-216.
5. Waring, M. J., Wakelin, L. P. G. Forty Years On. In *DNA and RNA Binders* (Demeunynck, M., Bailly, C., and Wilson, W. D., eds) Wiley-VCH: Weinheim, Germany, **2003**; 1, 1-17.
6. Lerman, L. S. Structural considerations in the interaction of DNA and acridines. *J. Mol. Biol.* **1961**, *3*, 18-30.
7. *Methods in Enzymology: Drug-Nucleic Acid Interactions*. Chaires J. B., Waring M. J. Academic Press: New York, **2001**.
8. *Sequence-specific DNA Binding Agents*. Waring M. Eds.; RSC Publishing: Cambridge, UK, **2006**.
9. Tsai, C. C., Jain, S. C., and Sobell, H. M. X-ray crystallographic visualization of drug-nucleic acid intercalative binding: structure of an ethidium-dinucleoside monophosphate crystalline complex, Ethidium: 5-iodouridylyl (3'-5') adenosine. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 628.

10. Wang, A. H., Nathans, J., van der Marel, G., van Boom, J. H., and Rich, A. Molecular structure of a double helical DNA fragment intercalator complex between deoxy CpG and a terpyridine platinum compound. *Nature* **1978**, 276, 471-474.
11. Shieh, H. S., Berman, H. M., Dabrow, M., and Neidle, S. The structure of drug-deoxydinucleoside phosphate complex; generalized conformational behavior of intercalation complexes with RNA and DNA fragments. *Nucleic Acids Res.* **1980**, 8, 85.
12. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R. E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. USA.* **1985**, 82, 1376-1380.
13. Coll, M., Frederick, C. A., Wang, A. H., and Rich, A. A bifurcated hydrogen-bonded conformation in the d (AT) base pairs of the DNA dodecamer d (CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, 84, 8385-8389.
14. Pelton, J. G., and Wemmer, D. E. Structural characterization of a 2: 1 distamycin Ad (CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, 86, 5723-5727.
15. Chen, X., Ramakrishnan, B., Rao, S. T., and Sundaralingam, M. Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex *Nat. Struct. Biol.* **1994**, 1, 169-175.
16. Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **2001**, 9, 2215-2235.
17. Dervan, P. B., and Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides *Curr. Opin. Struct. Biol.* **2003**, 13, 284-299. (For a cyclic polyamide crystal structure containing the  $\alpha$ -amino turn see Chapter 4 of this thesis.)
18. Rees, D. A. The future of biological X-ray analysis. In *Physical Biology: From Atoms to Medicine*. Zewail, A., Eds.; Imperial College Press: London, **2008**, 145-164.
19. Bloomfield, V. A., Crothers, D. M., Tinoco, I. *Nucleic Acids: Structure, Properties and Functions*. University Science Books: Sausalito, CA, **2000**
20. Neidle, S. *Nucleic Acid Structure and Recognition*. Oxford University Press: London, **2002**
21. Dickerson, R. E. *Present at the Flood: How Structural Molecular Biology Came About*. Sinauer Associates, Inc: Sunderland, MA, **2005**
22. Watson, J. D., and Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, 171, 737-738.
23. Miller, A., Tanner, J. *Essentials of Chemical Biology*. John Wiley & Sons, Ltd: West Sussex, England, **2008**
24. Arnott, S. Historical article: DNA polymorphism and the early history of the double helix. *Trends Biochem. Sci.* **2006**, 31, 349-354.
25. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, 389, 251-260.
26. Richmond, T. J., and Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **2003**, 423, 145-150.

27. Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **2005**, *436*, 138-141.
28. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., and others, O. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860-921.
29. Venter, J. C., et al. The sequence of the human genome. *Science* **2001**, *291*, 1304-1351.
30. International Human Genome Sequencing Consortium Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931-945.
31. Rice, P. A., Correll, C. C. *Protein-nucleic Acid Interactions: Structural Biology*. Rice P. A., Correll C. C. RSC Publishing: Cambridge, UK, **2008**
32. All ab initio calculations reported here were performed using HF/3-21G\* as implemented in the Gamess program on structures whose coordinates correspond to those of the crystal structure. Electrostatic potential surfaces were generated by mapping the electrostatic potentials onto surfaces of molecular electron density (0.002 electron/Å) and color-coding, using the Chimera program. The molecular electrostatic potential energy values range from -25 kcal/mol for values of negative potential (red) to +25 kcal/mol for values of positive potential (blue). This range was chosen to emphasize the variations in the aromatic region and some regions of the electrostatic potential associated with heteroatoms may lie beyond the ±25 kcal/mol range. 3-21G\* basis set: Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., Defrees, D. J., and Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The Journal of Chemical Physics*. **1982**, *77*, 3654-3665. (b) Hariharan, P. C., and Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*. **1973**, *28*, 213-222. Gamess program: Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., and Su, S. General atomic and molecular electronic structure system. *J. Comput. Chem*. **1993**, *14*, 1347-1363. UCSF Chimera: Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem*. **2004**, *25*, 1605-1612.
33. Wolberger, C. Multiprotein-DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 29-56.
34. Naar, A. M., Lemon, B. D., and Tjian, R. Transcriptional coactivator complexes. *Annu. Rev. Biochem.* **2001**, *70*, 475-501.
35. Panne, D., Maniatis, T., and Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **2007**, *129*, 1111-1123.
36. Seeman, N. C., Rosenberg, J. M., and Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 804.
37. Mrksich, M., Parks, M. E., and Dervan, P. B. Hairpin peptide motif. A new class of oligopeptides for sequence-specific recognition in the minor groove of double-helical DNA. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.
38. Herman, D. M., Baird, E. E., and Dervan, P. B. Stereochemical Control of the DNA Binding Affinity, Sequence Specificity, and Orientation Preference of Chiral Hairpin Polyamides in the

Minor Groove *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.

39. Dose, C., Farkas, M. E., Chenoweth, D. M., and Dervan, P. B. Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit. *J. Am. Chem. Soc.* **2008**, *130*, 6859-6866. For initial investigations into the (R)- and (S)- $\beta$ -amino-GABA turns and stereospecific synthetic routes, see Scott Carter's Thesis (reference 56 below).

40. Trauger, J. W., Baird, E. E., and Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **1996**, *382*, 559-561.

41. Herman, D. M., Turner, J. M., Baird, E. E., and Dervan, P. B. Cycle polyamide motif for recognition of the minor groove of DNA *J. Am. Chem. Soc.* **1999**, *121*, 1121-1129.

42. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., and Dervan, P. B. Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 7586.

43. Hsu, C. F., Phillips, J. W., Trauger, J. W., Farkas, M. E., Belitsky, J. M., Heckel, A., Olenyuk, B. Z., Puckett, J. W., Wang, C. C. C., and Dervan, P. B. Completion of a programmable DNA-binding small molecule library. *Tetrahedron* **2007**, *63*, 6146-6151.

44. Pilch, D. S., Poklar, N., Gelfand, C. A., Law, S. M., Breslauer, K. J., Baird, E. E., and Dervan, P. B. Binding of a hairpin polyamide in the minor groove of DNA: Sequence-specific enthalpic discrimination. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8306-8311.

45. Crothers, D. M., and Fried, M. Transmission of long-range effects in DNA. *Cold Spring Harb. Symp. Quant. Biol.* **1983**, *47*, 263-269.

46. Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. Structural basis for G•C recognition in the DNA minor groove. *Nat. Struct. Biol.* **1998**, *5*, 104-109.

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

48. Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. Structural effects of DNA sequence on TA recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **2000**, *295*, 557-567.

49. Greenberg, W. A., Baird, E. E., and Dervan, P. B. A Comparison of H-Pin and Hairpin Polyamide Motifs for the Recognition of the Minor Groove of DNA

50. Olenyuk, B., Jitianu, C., and Dervan, P. B. Parallel synthesis of H-pin polyamides by alkene metathesis on solid phase. *J. Am. Chem. Soc.* **2003**, *125*, 4741-4751.

51. Heckel, A., and Dervan, P. B. U-pin polyamide motif for recognition of the DNA minor groove. *Chem. Eur. J.* **2003**, *9*, 1-14.

52. Cho, J., Parks, M. E., and Dervan, P. B. Cyclic polyamides for recognition in the minor groove of DNA. *Proc. Natl. Acad. Sci. USA.* **1995**, *92*, 10389-10392.

53. Melander, C., Herman, D. M., and Dervan, P. B. Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif. *Chemistry* **2000**, *6*, 4487-4497.

54. Baliga, R., Baird, E. E., Herman, D. M., Melander, C., Dervan, P. B., and Crothers, D. M. Kinetic consequences of covalent linkage of DNA binding polyamides. *Biochemistry* **2001**, *40*,

3-8.

55. Zhang, Q., Dwyer, T. J., Tsui, V., Case, D. A., Cho, J., Dervan, P. B., and Wemmer, D. E. NMR structure of a cyclic polyamide-DNA complex. *J. Am. Chem. Soc.* **2004**, *126*, 7958-7966.

56. Carter, S. R. *Sequence-specific minor groove binding polyamides: DNA recognition and applications*. California Institute of Technology, **1998**.

57. Urbach, A. R., Love, J. J., Ross, S. A., and Dervan, P. B. Structure of a beta-alanine-linked polyamide bound to a full helical turn of purine tract DNA in the 1: 1 motif. *J. Mol. Biol.* **2002**, *320*, 55-71.

58. Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* **2003**, *326*, 371-380.

59. Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Molecular recognition of the nucleosomal "supergroove". *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6864-6869.

60. Berman, H. M. et al. "The Protein Data Bank" *Nucleic Acids Res.* **2000**, *28*, 235-242.

## Chapter 2: Solution-Phase Synthesis of Pyrrole–Imidazole Polyamides

*The text of this chapter was taken in part from a manuscript coauthored with Daniel A. Harki, and Peter B. Dervan\* (Caltech)*

(Chenoweth, D.M., Harki, D.A., Dervan, P. B. *J. Am. Chem. Soc.* **2009** In Press.)

**Abstract**

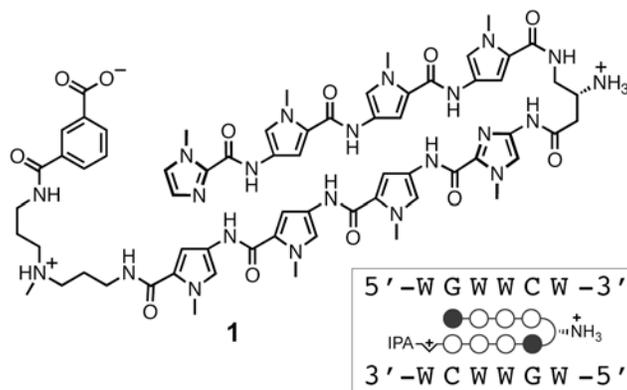
Pyrrole–imidazole polyamides are DNA-binding molecules that are programmable for a large repertoire of DNA sequences. Typical syntheses of this class of heterocyclic oligomers rely on solid-phase methods. Solid-phase methodologies offer rapid assembly on a micromole scale sufficient for biophysical characterizations and cell culture studies. In order to produce gram-scale quantities necessary for efficacy studies in animals, polyamides must be readily synthesized in solution. An 8-ring hairpin polyamide **1**, which targets the DNA sequence 5'-WGWWCW-3', was chosen for our synthesis studies as this oligomer exhibits androgen receptor antagonism in cell culture models of prostate cancer. A convergent solution-phase synthesis of **1** from a small set of commercially available building blocks is presented which highlights principles for preparing gram quantities of pyrrole–imidazole oligomers with minimal chromatography.

## 2.1 Introduction

Pyrrole–imidazole polyamides are a class of small molecules that bind the minor groove of DNA sequence-specifically.<sup>1,2</sup> Encoded by side-by-side arrangements of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamide monomers, Im/Py pairs distinguish G•C from C•G base pairs, whereas Py/Py pairs are degenerate for T•A and A•T.<sup>3</sup> Hairpin Py-Im polyamides have been shown to bind a broad repertoire of DNA sequences,<sup>4</sup> permeate cell membranes and traffic to the nucleus,<sup>5</sup> access chromatin,<sup>6</sup> and disrupt protein–DNA interfaces.<sup>2</sup> Hairpin polyamide inhibition of transcription factor–DNA binding of HIF-1 $\alpha$ ,<sup>7</sup> androgen receptor (AR),<sup>8</sup> and AP-1<sup>9</sup> has been exploited for controlling expression of medically relevant genes such as VEGF, PSA, TGF- $\beta$ 1, and LOX-1 in cell culture experiments.

An underpinning to transition polyamide studies from cell culture to small animal disease models is the ability to synthesize Py-Im polyamides on gram-scale. Over the years advances in polyamide solid-phase synthesis have been reported, including Boc- and Fmoc-based approaches from our laboratories and others.<sup>10</sup> Solid-phase methodologies offer many advantages for milligram-scale polyamide syntheses, including rapid and reliable amino acid couplings and facile purifications owing to immobilization of the polyamide oligomer on a solid support. However, these techniques intrinsically limit the scale of synthesis. Conversely, efficient gram-scale solution-phase methods for polyamide synthesis that avoid arduous chromatographic purifications and employ commercially available Py-Im amino acid building blocks as reagents are less well developed. Remarkably, solution-phase synthesis of hairpin polyamides was the standard in our laboratory prior to the development of solid-phase methodologies,<sup>11a</sup> and many variations on this theme have been published.<sup>11</sup> However, laborious chromatographic purifications and modest reaction yields are commonplace. Therefore, we sought to develop a general solution-phase polyamide synthesis method that would allow access to gram quantities of material in high yield with minimal chromatography.

We report a proof-of-principle study demonstrating that hairpin Py-Im polyamides can be synthesized in solution from a small set of building blocks on large scale with minimal use of chromatography. This method involves Boc-protected dimers, trimers, and tetramers of heterocycles suitable for convergent syntheses. By exploiting differences in the physical solubility properties of starting materials versus products, a solution-phase synthesis of an 8-ring hairpin Py-Im polyamide **1** (Figure 2.1) has been achieved. Notably, our synthesis permits core polyamide **2** (Figure 2.2) to be prepared without a single chromatographic purification, thereby providing large quantities



**Figure 2.1** Structure of Py-Im hairpin polyamide **1** targeting the DNA sequence 5'-WGWWCW-3' and its ball-and-stick representation.

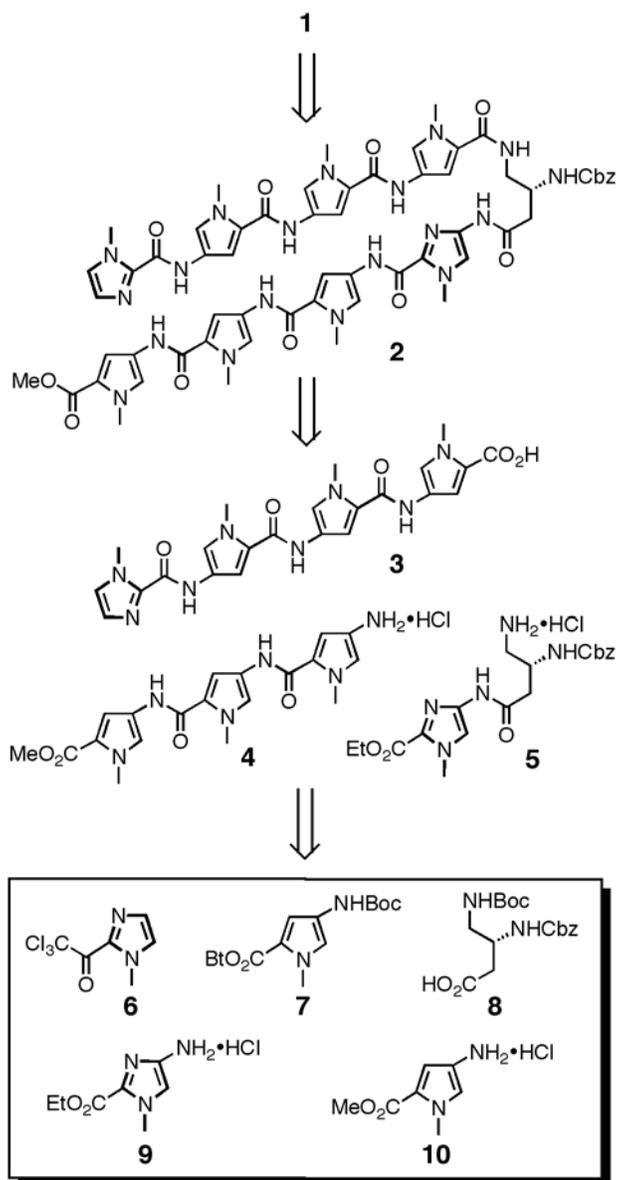
of **2** for subsequent modification at the C-terminus such as **1**. Py-Im polyamide **1**, which targets the DNA sequence 5'-WGWWCW-3', was selected for our studies because it antagonizes AR binding to androgen response elements (ARE) in gene promoters and regulates a subset of AR-driven genes, such as PSA.<sup>5g</sup> The regulation of aberrant AR-activated gene expression in prostate cancer is a promising strategy for developing novel therapeutics.<sup>8</sup> This

biological activity, coupled with our desire to conduct small animal efficacy experiments with **1**, renders this polyamide an ideal candidate for scale-up and optimization studies. In addition, we discuss unifying principles for planning solution-phase polyamide syntheses of different Py-Im arrangements.

## 2.2 Results and Discussion

Our retrosynthetic approach for the preparation of an 8-ring hairpin polyamide, ImPyPyPy-(*R*)<sup>β-H<sub>2</sub>N</sup>γ-ImPyPyPy-(+)-IPA (**1**), is shown in Figure 2.2. Sequential couplings of ImPyPyPy tetramer **3** to turn moiety **5**, followed by ester saponification and coupling to PyPyPy trimer **4**, afford polyamide **2** in a convergent manner. Advanced intermediates **3–5** were prepared from building blocks **6–10**,<sup>12</sup> which have been previously synthesized by our laboratory and others.<sup>10a,10c,13</sup> The cornerstone of our synthesis strategy capitalizes on the disparate physical properties of starting materials versus products, which permit purification of each intermediate to be achieved by combinations of precipitation, trituration, and crystallization. Such details are described below.

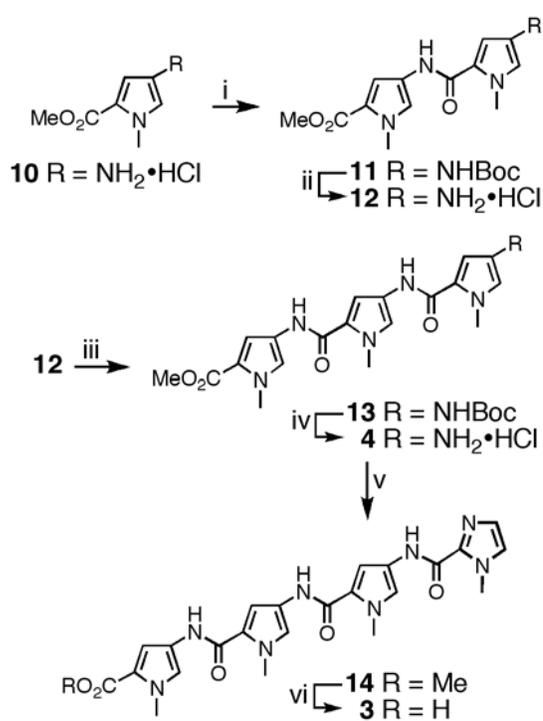
The synthesis of pyrrole trimer **4** begins with pyrrole amine salt **10** (Scheme 2.1). Amide coupling of **10** with activated pyrrole monomer **7** affords dimer **11** in 93% yield. The utilization of a small excess of **10** relative to **7** drives the reaction to completion, and residual **10** is readily separated from **11** following precipitation in water and aqueous washing of the residual solid **11**. Reaction of dimer **11** with anhydrous HCl in diethyl ether removes the carbamate protecting group and facilitates precipitation of **12** as the HCl salt during the course of the reaction. Isolation of solid **12** by filtration, followed by washing of the residual solid with excess Et<sub>2</sub>O, provides the amine



**Figure 2.2** Retrosynthetic strategy for the convergent solution-phase synthesis of polyamide **1**. **7–10** are commercially available building blocks. Boc = *tert*-butyl carbamate, Bt = benzotriazole, Cbz = benzyl carbamate.

HCl salt in 99% yield. By exploiting the aqueous solubility of **10** versus insolubility of **11**, PyPy dimer **11** is easily purified from a small excess of **10** by precipitation, whereas deprotected **12** is separable from Boc-protected **11** by virtue of the Et<sub>2</sub>O solubility of **11** versus insolubility of **12**. This reaction sequence highlights our synthesis strategy: exploiting the different solubility profiles of reactants versus products for chromatography-free purifications. Accordingly, pyrrole trimer **4** was obtained from dimer **12** by coupling with **7**, followed by acidic deprotection to yield **4** in 95% yield (two steps) from dimer **12**. The ImPyPyPy tetramer **3** was synthesized in two steps from trimer **4**. 1-Methyl-2-trichloroacetylimidazole (**6**), prepared in one step from *N*-methylimidazole,<sup>10c,13</sup> was allowed to react with a small excess of **4** to deliver tetramer **14** in 83% yield following precipitation in H<sub>2</sub>O, trituration with Et<sub>2</sub>O, and drying *in vacuo*. Saponification of **14** with aqueous NaOH in 1,4-dioxane, followed by neutralization with aqueous HCl, precipitation, and Et<sub>2</sub>O trituration, afforded tetramer **3** in 77% overall yield from trimer **4**.

The Im-turn fragment **5** was synthesized in two steps from the Im•HCl salt monomer<sup>10a</sup> **9** by coupling to PyBOP-activated (*R*)-3,4-Cbz-Dbu(Boc)-OH (**8**) yielding protected dimer **15** in 95% yield (Scheme 2.2). The utilization of a small excess of **9** drives the coupling reaction to completion and is easily separated in the aqueous wash step. Removal of the carbamate protecting



**Scheme 2.1** Preparation of **3** and **4**. Reagents and Conditions: (i) DMF, DIEA, **7**, 23 °C, 8 h, 93%; (ii) 2.0 M HCl in Et<sub>2</sub>O, 23 °C, 18 h, 99%; (iii) DMF, DIEA, **7**, 23 °C, 8 h, 96%; (iv) 4.0 M HCl in 1,4-dioxane, 23 °C, 18 h, 99%; (v) DMF, DIEA, **6**, 23 °C, 2 h, 83%; (vi) NaOH (aq), 1,4-dioxane, 42 °C, 2 h, 93%.

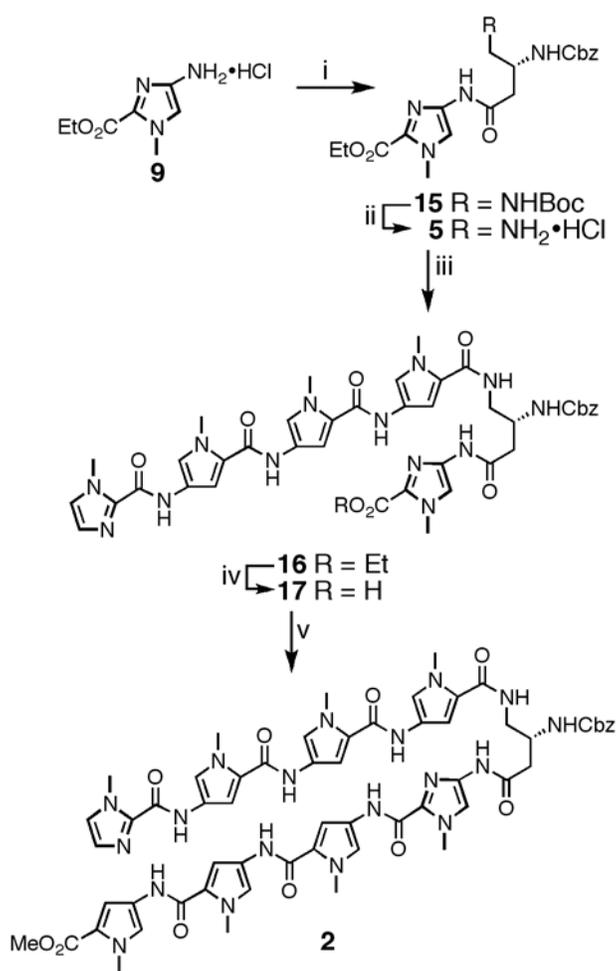
group with anhydrous HCl in 1,4-dioxane yielded the final synthon for our studies, imidazole-turn dimer **5**, in quantitative yield following filtration and washing of the residual salt. With compound **5** in hand, the assembly of core polyamide **2** was initiated. PyBOP-mediated coupling of tetramer **3** to a small excess of water-soluble Im-turn dimer **5** yielded the advanced intermediate **16** in 97% yield. Saponification of **16** to acid **17**, followed by amide coupling with an excess of water-soluble trimer **4**, delivered core Py-Im hairpin polyamide **2** in 88% yield for the two steps. Consistent with the previously discussed strategy for synthesizing intermediates **3–5**, the differences in solubility

of reactants versus products were exploited to isolate pure material by precipitation, washing, and trituration. In most cases a low-boiling-point solvent was employed in the final trituration step to facilitate efficient solvent removal *in vacuo*.

Core polyamide **2** was synthesized without a single chromatographic purification in high

overall purity, as depicted by the analytical HPLC analysis of **2** shown in Figure 2.3. Multigram quantities of **2** have been readily synthesized by this method, providing a stockpile of material for elaboration at the C-terminus into discrete polyamide conjugates, such as **1**.

Py-Im polyamide **1** was synthesized in solution from advanced core **2** by coupling the preassembled C-terminal tail moiety **21** with saponified core **22**. This convergent approach begins by preparing Boc-protected C-terminus moiety **20** (Scheme 2.3). PyBOP-mediated coupling of mono-Boc-protected triamine linker **18**<sup>14</sup> with monobenzyl-protected isophthalic acid **19**<sup>15</sup> afforded Boc-protected **20** in 98% yield. Deprotection of **20** with anhydrous CF<sub>3</sub>CO<sub>2</sub>H in dichloromethane (1:1) yielded amine **21**, which was used immediately following concentration under high vacuum. Saponification of core polyamide **2** with aqueous NaOH in 1,4-dioxane at 23 °C yielded 8-ring acid **22** in 89% yield (Scheme 2.4). The transformation of **2** to **22** proved somewhat difficult

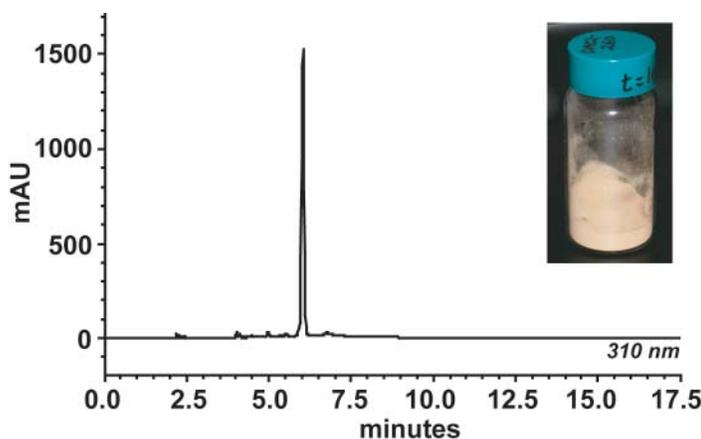


**Scheme 2.2** Preparation of **5** and assembly of core polyamide **2**. Reagents and Conditions: (i) DMF, DIEA, PyBOP,  $(R)$ -3,4-Cbz-Dbu(Boc)-OH (**8**), 23 °C, 8 h, 95%; (ii) HCl in 1,4-dioxane, 23 °C, 16 h, 99%; (iii) DMF, DIEA, PyBOP, **3**, 23 °C, 2 h, 97%; (iv) KOH (aq), MeOH, 1,4-dioxane, 42 °C, 2 h, 92%; (i) DMF, DIEA, PyBOP, **4**, 23 °C, 10 h, 96%.

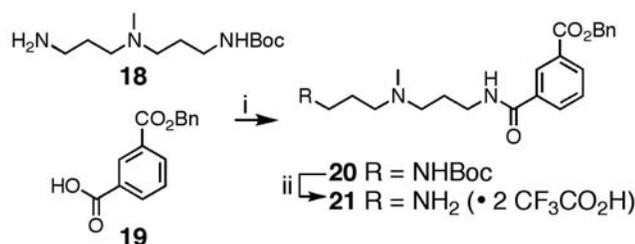
in early studies due to formation of an unidentified side product. Avoiding reaction temperatures above 23 °C suppresses most of the byproduct formation, whereas a screen of aqueous bases commonly used for ester saponification (KOH and LiOH) failed to identify a better reagent. Coupling of residual acid **22** with freshly prepared **21** delivered the penultimate oligomer **23** in 87% yield. Unfortunately, crude oligomer **23** could not be satisfactorily purified by our standard method and required chromatography on silica gel to achieve pure material. Hundreds of milligrams of **23** have been prepared by this method in a single reaction sequence. Global deprotection of **23** by hydrogenation (Pd/C, ~1 atm H<sub>2</sub>) at 23 °C for 48 h yields Py-Im polyamide **1** in 81% yield. Final product **1** can be separated from residual catalyst by solid-phase extraction and then purified by preparative reverse-phase HPLC.

With multi-milligram quantities of **1** on hand, we investigated in detail the UV properties of Py-Im polyamide **1** in a variety

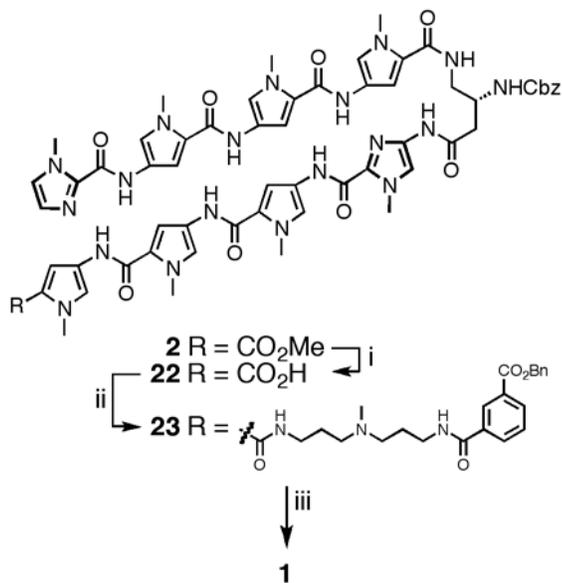
of laboratory and biologically relevant solvents. As shown in Figure 2.4, a strong solvent influence on the molar extinction coefficient of **1** is observed as the amount of organic cosolvent is increased. For example, an extinction coefficient ( $\epsilon$ , M<sup>-1</sup>cm<sup>-1</sup>) of 26500 was measured for **1** in distilled and deionized H<sub>2</sub>O, whereas this value doubled to 54800 in 50% acetonitrile in aqueous CF<sub>3</sub>CO<sub>2</sub>H (0.1% v/v CF<sub>3</sub>CO<sub>2</sub>H), a widely utilized laboratory solvent for purifying and quantifying peptides. A stock solution frequently encountered for preparing biological samples, 10% DMSO in DEPC-treated H<sub>2</sub>O yielded an intermediary value of 40900 M<sup>-1</sup>cm<sup>-1</sup>. Hence, care must be taken to consider the solvent



**Figure 2.3** Analysis of polyamide **2** purity by analytical HPLC. Wavelength shown is 310 nm.



**Scheme 2.3** Preparation of **21**. Reagents and Conditions: (i) DMF, DIEA, PyBOP, 23 °C, 3 h, 98%; (ii)  $\text{CF}_3\text{CO}_2\text{H}$ ,  $\text{CH}_2\text{Cl}_2$ , 23 °C, 20 min, concentration *in vacuo* and used crude.



**Scheme 2.4** Final steps for the synthesis of Py-Im polyamide **1**. Reagents and Conditions: (i) NaOH (aq), 1,4-dioxane, 23 °C, 11 h, 89%; (ii) DIEA, DMF, PyBOP, **21**, 23 °C, 12 h, 87%; (iii) DMF, Pd/C,  $\text{H}_2$  (1 atm), 23 °C, 48 h, 81%.

system utilized when performing UV spectroscopy to determine polyamide concentrations.

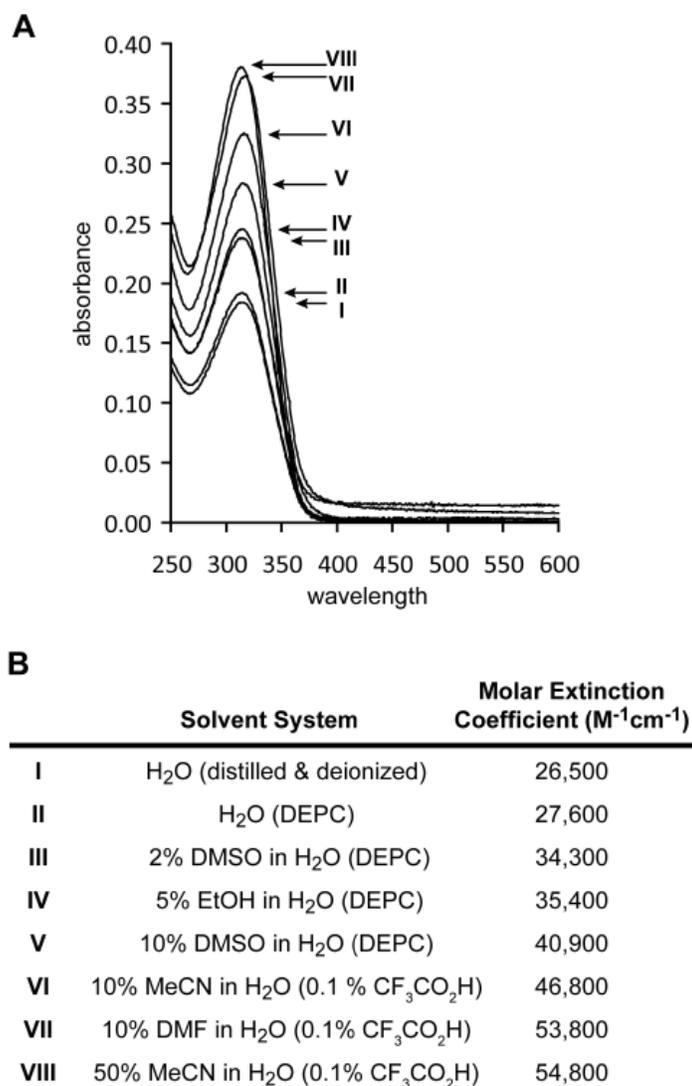
### 2.3 Conclusion

A solution-phase synthesis of Py-Im hairpin polyamide **1** is presented, highlighting unifying principles for the preparation of related polyamides. A convergent synthesis was developed, requiring no chromatographic purifications to arrive at core 8-ring polyamide **2** on multi-gram scale. Final elaboration of the C-terminus affords AR polyamide antagonist **1** in high yield. The synthetic methodology permits gram-scale synthesis of Py-Im polyamides, a minimum next step as we transition these small molecules to animal models for biological efficacy.

### 2.4 Experimental Section

#### 2.4.1 General

Chemicals were purchased from Sigma-Aldrich and were used without further purification. (*R*)-3,4-Cbz-Dbu(Boc)-OH was purchased from Senn Chemicals AG (code number 44159). Bulk grade solvents were from Fisher Scientific. Centrifugation



**Figure 2.4** UV properties of polyamide **1** in solvent systems I–VIII. (A) UV traces of I–VIII from 250–600 nm and (B) tabular form of data. Molar extinction coefficients were calculated from the  $\lambda_{max}$  for each individual system, which ranged from 313 to 317 nm. DEPC = diethylpyrocarbonate treated H<sub>2</sub>O.

was performed in a Beckman Coulter benchtop centrifuge (Allegra 21R) equipped with a Beckman swing-out rotor (model S4180). Analytical HPLC analysis was conducted on a Beckman Gold instrument equipped with a Phenomenex Gemini analytical column (250 × 4.6 mm, 5  $\mu$ m) and a diode array detector, and the mobile phase consisted of a gradient of acetonitrile (MeCN) in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H. Preparative HPLC was performed on an Agilent 1200 system equipped with a solvent degasser, a diode array detector, and a Phenomenex Gemini column (250 × 21.2 mm, 5  $\mu$ m). A gradient of MeCN in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H was utilized as the mobile phase. UV–Vis measurements were made on a Hewlett-Packard diode array spectrophotometer (model 8452 A). NMR spectroscopy was performed on a Varian instrument operating at 499.8 MHz (for <sup>1</sup>H) or 125.7 MHz (for <sup>13</sup>C) at ambient temperature. All NMR analyses were performed in DMSO-*d*<sub>6</sub>, and chemical shifts are reported in parts per million relative to the internal solvent peak referenced to 2.49 (for <sup>1</sup>H) or 39.5 (for <sup>13</sup>C). High-resolution mass spectrometry (HRMS) was recorded in positive-ion mode by fast-atom bombardment (FAB<sup>+</sup>) on a JEOL JMS-600H instrument or by electrospray ionization (ESI<sup>+</sup>) on a Waters Acquity UPLC-LCT Premiere XE TOF-MS system.

was performed in a Beckman Coulter benchtop centrifuge (Allegra 21R) equipped with a Beckman swing-out rotor (model S4180). Analytical HPLC analysis was conducted on a Beckman Gold instrument equipped with a Phenomenex Gemini analytical column (250 × 4.6 mm, 5  $\mu$ m) and a diode array detector, and the mobile phase consisted of a gradient of acetonitrile (MeCN) in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H. Preparative HPLC was performed on an Agilent 1200 system equipped with a solvent degasser, a diode array detector, and a Phenomenex Gemini column (250 × 21.2 mm, 5  $\mu$ m). A gradient of MeCN in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H was utilized as the mobile phase. UV–Vis measurements were made on a Hewlett-Packard diode array spectrophotometer (model 8452 A). NMR spectroscopy was performed on a Varian instrument operating at 499.8 MHz (for <sup>1</sup>H) or 125.7 MHz (for <sup>13</sup>C) at ambient temperature. All NMR analyses were

#### 2.4.2 HCl•H<sub>2</sub>N-Py-CO<sub>2</sub>Me (**10**)

Prepared as previously described.<sup>10a</sup> <sup>1</sup>H NMR: δ 10.09 (br s, 3H), 7.25 (d, *J* = 2.0 Hz, 1H), 6.80 (d, *J* = 2.2 Hz, 1H), 3.85 (s, 3H), 3.74 (s, 3H); <sup>13</sup>C NMR: δ 160.2, 123.7, 120.8, 113.9, 111.4, 51.3, 36.6; HRMS (FAB<sup>+</sup>) calc'd for C<sub>7</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub> [M+H]<sup>+</sup> 155.0821, found 155.0847.

#### 2.4.3 BocHN-PyPy-CO<sub>2</sub>Me (**11**)

A solution of BocHN-Py-OBt **7** (6.16 g, 17.2 mmol) and HCl•H<sub>2</sub>N-Py-CO<sub>2</sub>Me **10** (3.61 g, 19.0 mmol) in DMF (39 mL) and DIEA (6 mL, 34.4 mmol) was stirred at 23 °C for 8 h. The solution was then added to distilled H<sub>2</sub>O (500 mL) preacidified with aqueous HCl (1 N, 300 mL, 300 mmol), yielding a precipitate that was isolated by centrifugation (~ 4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (80 mL) and collected by centrifugation (repeated 2X). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. Drying of the light-brown solid *in vacuo* yielded dimer **11** (6.03 g, 93%). <sup>1</sup>H NMR: δ 9.84 (s, 1H), 9.10 (s, 1H), 7.44 (d, *J* = 1.7 Hz, 1H), 6.88 (m, 2H), 6.82 (s, 1H), 3.82 (s, 3H), 3.79 (s, 3H), 3.72 (s, 3H), 1.44 (s, 9H); <sup>13</sup>C NMR: δ 160.8, 158.4, 152.8, 123.0, 122.6, 122.4, 120.7, 118.4, 117.1, 108.3, 103.8, 78.3, 50.9, 36.1, 36.0, 28.2; HRMS (FAB<sup>+</sup>) calc'd for C<sub>18</sub>H<sub>25</sub>N<sub>4</sub>O<sub>5</sub> [M+H]<sup>+</sup> 377.1825, found 377.1835.

#### 2.4.4 HCl•H<sub>2</sub>N-PyPy-CO<sub>2</sub>Me (**12**)

Dimer **11** (4.0 g, 10.6 mmol) in a solution of anhydrous HCl in Et<sub>2</sub>O (2.0 M, 400 mL) was stirred at 23 °C for 18 h. The mixture was then diluted with 400 mL of anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield dimer **12** as a tan solid (3.3 g, 99%). <sup>1</sup>H NMR: δ 10.12 (s, 1H), 10.07 (br s, 3H), 7.46 (d, *J* = 2.0 Hz, 1H), 7.10 (d, *J* = 2.0 Hz, 1H), 7.00 (d, *J* = 2.0 Hz, 1H), 6.91 (d, *J* = 2.0 Hz, 1H), 3.87 (s, 3H), 3.83 (s, 3H), 3.72 (s, 3H); <sup>13</sup>C NMR: δ 160.8, 157.7, 124.6, 122.6, 121.7, 120.8, 118.6, 113.1, 108.4, 107.2, 51.0, 36.6, 36.2; HRMS (FAB<sup>+</sup>) calc'd for C<sub>13</sub>H<sub>17</sub>N<sub>4</sub>O<sub>3</sub> [M+H]<sup>+</sup> 277.1301, found 277.1292.

#### 2.4.5 BocHN-PyPyPy-CO<sub>2</sub>Me (**13**)

A solution of BocHN-Py-OBt **7** (3.1 g, 8.7 mmol) and dimer **12** (3.0 g, 9.59 mmol) in DMF (20 mL) and DIEA (3 mL, 17.4 mmol) was stirred at 23 °C for 8 h. The solution was then added to distilled H<sub>2</sub>O (250 mL) preacidified with aqueous HCl (1 N, 150 mL, 150 mmol), yielding a precipitate that

was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (40 mL) and collected by centrifugation (repeated 2x). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. Drying of the light-brown solid *in vacuo* yielded trimer **13** (4.17 g, 96%). <sup>1</sup>H NMR: δ 9.91 (s, 1H), 9.86 (s, 1H), 9.09 (s, 1H), 7.46 (d, *J* = 2.0 Hz, 1H), 7.21 (d, *J* = 1.7 Hz, 1H), 7.05 (d, *J* = 1.5 Hz, 1H), 6.89 (m, 2H), 6.83 (s, 1H), 3.83 (s, 6H), 3.80 (s, 3H), 3.73 (s, 3H), 1.45 (s, 9H); <sup>13</sup>C NMR: δ 160.8, 158.5, 158.4, 152.8, 123.0, 122.8, 122.4, 122.30, 122.29, 120.7, 118.48, 118.47, 117.0, 108.3, 104.8, 103.8, 78.3, 50.9, 36.2, 36.05, 36.04, 28.2; HRMS (FAB<sup>+</sup>) calc'd for C<sub>24</sub>H<sub>30</sub>N<sub>6</sub>O<sub>6</sub> [M]<sup>+</sup> 498.2227, found 498.2233.

#### 2.4.6 HCl•H<sub>2</sub>N-PyPyPy-CO<sub>2</sub>Me (**4**)

Trimer **13** (4.0 g, 8.02 mmol) in a solution of anhydrous HCl in 1,4-dioxane (4.0 M, 300 mL) was stirred at 23 °C for 18 h. The mixture was then diluted with 600 mL of anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield trimer **4** as a brown-orange solid (3.45 g, 99%). <sup>1</sup>H NMR: δ 10.16 (s, 3H), 10.13 (s, 1H), 9.97 (s, 1H), 7.46 (d, *J* = 2.0 Hz, 1H), 7.25 (d, *J* = 1.7 Hz, 1H), 7.11 (d, *J* = 2.0 Hz, 1H), 7.08 (d, *J* = 2.0 Hz, 1H), 7.02 (d, *J* = 2.0 Hz, 1H), 6.91 (d, *J* = 2.0 Hz, 1H), 3.88 (s, 3H), 3.84 (s, 3H), 3.82 (s, 3H), 3.72 (s, 3H); <sup>13</sup>C NMR: δ 160.8, 158.4, 157.7, 124.8, 123.0, 122.6, 121.9, 121.6, 120.8, 118.7, 118.5, 113.0, 108.4, 107.2, 104.8, 51.0, 36.6, 36.2, 36.1; HRMS (FAB<sup>+</sup>) calc'd for C<sub>19</sub>H<sub>22</sub>N<sub>6</sub>O<sub>4</sub> [M]<sup>+</sup> 398.1702, found 398.1685.

#### 2.4.7 ImPyPyPy-CO<sub>2</sub>Me (**14**)

A solution of trimer **4** (1.019 g, 2.34 mmol) and 1-methyl-2-trichloroacetylimidazole (**6**)<sup>10c,13</sup> (478 mg, 2.10 mmol) in DMF (4.5 mL) and DIEA (910 μL, 5.22 mmol) was stirred at 23 °C for 2 h. The solution was then added to distilled H<sub>2</sub>O (40 mL) pre-acidified with aqueous HCl (1N, 910 μL, 0.91 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (40 mL) and collected by centrifugation (repeated 1X). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield tetramer **14** as a light brown solid (886 mg, 83%). <sup>1</sup>H NMR: δ 10.68 (s, 1H), 10.00 (s, 1H), 9.94 (s, 1H), 7.48 (s, 1H), 7.47 (d, *J* = 2.0 Hz, 1H), 7.31 (d, *J* = 1.7 Hz, 1H), 7.24 (d, *J* = 1.7 Hz, 1H), 7.18 (s, 1H), 7.17 (d, *J* = 2.0 Hz, 1H), 7.08 (d, *J* = 2.0 Hz, 1H), 6.91 (d, *J* = 1.7 Hz, 1H), 4.01 (s, 3H), 3.86 (s,

3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.73 (s, 3H);  $^{13}\text{C}$  NMR:  $\delta$  160.8, 158.5, 158.4, 155.0, 138.2, 126.4, 125.6, 123.1, 123.0, 122.5, 122.2, 121.2, 120.7, 118.8, 118.6, 118.5, 108.3, 104.9, 104.8, 50.9, 36.18, 36.17, 36.1, 35.4; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{24}\text{H}_{27}\text{N}_8\text{O}_5$  [M+H]<sup>+</sup> 507.2104, found 507.2116.

#### 2.4.8 *ImPyPyPy-CO<sub>2</sub>H* (**3**)

A solution of tetramer **14** (804 mg, 1.59 mmol) in 1,4-dioxane (8 mL) and aqueous NaOH (1N, 8.0 mL, 8.00 mmol) was stirred at 42 °C for 2 h. The solution was then added to distilled H<sub>2</sub>O (40 mL) pre-acidified with aqueous HCl (1N, 8 mL, 8.00 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (40 mL) and collected by centrifugation (repeated 1X). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield tetramer **3** as a brown solid (728 mg, 93%).  $^1\text{H}$  NMR:  $\delta$  10.45 (s, 1H), 9.97 (s, 1H), 9.91 (s, 1H), 7.43 (s, 1H), 7.38 (s, 1H), 7.29 (s, 1H), 7.24 (s, 1H), 7.17 (s, 1H), 7.07 (s, 1H), 7.03 (s, 1H), 6.85 (s, 1H), 3.99 (s, 3H), 3.85 (s, 3H), 3.84 (s, 3H), 3.82 (s, 3H);  $^{13}\text{C}$  NMR:  $\delta$  162.0, 158.48, 158.46, 156.1, 138.8, 127.0, 126.4, 123.0, 122.7, 122.6, 122.2, 121.5, 120.3, 119.5, 118.7, 118.5, 108.4, 105.0, 104.8, 36.14, 36.07, 35.1; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{23}\text{H}_{25}\text{N}_8\text{O}_5$  [M+H]<sup>+</sup> 493.1948, found 493.1952.

#### 2.4.9 *BocHN-(R)<sup>β</sup>-CbzHN $\gamma$ -Im-CO<sub>2</sub>Et* (**15**)

A solution of (*R*)-3,4-Cbz-Dbu(Boc)-OH **8** (1.03 g, 2.93 mmol) and PyBOP (1.83 g, 3.51 mmol) in DMF (12 mL) and DIEA (1.5 mL, 8.8 mmol) was stirred at 23 °C for 22 min. The solution was then added to solid (powdered) HCl•H<sub>2</sub>N-Im-CO<sub>2</sub>Et **9** (850 mg, 3.15 mmol) and stirred at 23 °C for 8 h. The solution was then added to distilled H<sub>2</sub>O (30 mL) pre-acidified with aqueous HCl (1N, 9 mL, 9 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (40 mL) and collected by centrifugation (repeated 2X). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. Drying of the brown solid *in vacuo* yielded dimer **15** (1.4 g, 95%).  $^1\text{H}$  NMR:  $\delta$  10.57 (s, 1H), 7.51 (s, 1H), 7.31-7.27 (m, 5H), 7.02 (d,  $J$  = 8.5 Hz, 1H), 6.79 (m, 1H), 4.97 (s, 2H), 4.25 (q,  $J$  = 7.2 Hz, 2H), 3.93 (m, 1H), 3.89 (s, 3H), 3.01 (m, 2H), 2.44-2.35 (m, 2H), 1.35 (s, 9H), 1.27 (t,  $J$  = 7.2 Hz, 3H);  $^{13}\text{C}$  NMR:  $\delta$  167.6, 158.4, 155.8, 155.4, 137.4, 137.1, 130.7, 128.2, 127.6, 127.5, 114.8, 77.7, 65.1, 60.5, 48.6, 43.5, 38.1, 35.4, 28.2, 14.0; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{24}\text{H}_{34}\text{N}_5\text{O}_7$

$[M+H]^+$  504.2458, found 504.2462.

#### 2.4.10 HCl•H<sub>2</sub>N-(R)<sup>β-CbzHN</sup>γ-Im-CO<sub>2</sub>Et (**5**)

Dimer **15** (500 mg, 0.993 mmol) in a solution of anhydrous HCl in 1,4-dioxane (4.0 M, 10 mL) and anhydrous Et<sub>2</sub>O (4 mL) was stirred at 23 °C for 16 h. The mixture was then diluted with 20 mL of anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield dimer **5** as a white solid (432 mg, 99%). <sup>1</sup>H NMR: δ 10.75 (s, 1H), 8.05 (m, 3H), 7.52 (s, 1H), 7.38 (d, *J* = 8.5 Hz, 1H), 7.34-7.28 (m, 5H), 5.01 (m, 2H), 4.25 (q, *J* = 7.1 Hz, 2H), 4.13 (m, 1H), 3.90 (s, 3H), 2.96-2.84 (m, 2H), 2.60-2.51 (m, 2H), 1.27 (t, *J* = 7.1 Hz, 3H); <sup>13</sup>C NMR: δ 166.8, 158.4, 155.7, 137.2, 136.8, 130.9, 128.3, 127.8, 127.7, 114.9, 65.5, 60.6, 46.6, 42.1, 38.2, 35.4, 14.0; HRMS (FAB<sup>+</sup>) calc'd for C<sub>19</sub>H<sub>26</sub>N<sub>5</sub>O<sub>5</sub>  $[M+H]^+$  404.1934, found 404.1928.

#### 2.4.11 ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-Im-CO<sub>2</sub>Et (**16**)

A solution of tetramer **3** (1.28 g, 2.59 mmol) and PyBOP (1.49 g, 2.86 mmol) in DMF (8 mL) and DIEA (1.8 mL, 10.4 mmol) was stirred at 23 °C for 20 min. The solution was then treated with solid (powdered) HCl•H<sub>2</sub>N-(R)<sup>β-CbzHN</sup>γ-Im-CO<sub>2</sub>Et **5** (1.2 g, 2.73 mmol) and stirred at 23 °C for 2 h. The solution was then added to distilled H<sub>2</sub>O (30 mL) pre-acidified with aqueous HCl (1 N, 20 mL, 20 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (40 mL) and collected by centrifugation (repeated 2x). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-Im-CO<sub>2</sub>Et **16** as a tan solid (2.2 g, 97%). <sup>1</sup>H NMR: δ 10.60 (s, 1H), 10.46 (s, 1H), 9.96 (s, 1H), 9.91 (s, 1H), 7.97 (t, *J* = 5.6 Hz, 1H), 7.52 (s, 1H), 7.39 (s, 1H), 7.29-7.26 (m, 6H), 7.24 (m, 1H), 7.18-7.14 (m, 3H), 7.04 (m, 2H), 6.90 (m, 1H), 4.98 (m, 2H), 4.24 (q, *J* = 7.1 Hz, 2H), 4.10 (m, 1H), 3.99 (s, 3H), 3.89 (s, 3H), 3.845 (s, 3H), 3.839 (s, 3H), 3.77 (s, 3H), 3.29 (m, 2H), 2.49 (m, 2H, obstructed by NMR solvent), 1.27 (t, *J* = 7.2 Hz, 3H); <sup>13</sup>C NMR: δ 167.7, 161.5, 158.49, 158.48, 158.45, 156.0, 155.6, 138.7, 137.4, 137.1, 130.8, 128.2, 127.7, 127.6, 126.8, 126.4, 123.0, 122.8, 122.7, 122.24, 122.17, 121.4, 118.7, 118.5, 118.0, 114.9, 105.0, 104.7, 104.4, 65.2, 60.5, 48.6, 42.1, 38.2, 36.13, 36.11, 36.0, 35.4, 35.2, 14.0; HRMS (FAB<sup>+</sup>) calc'd for C<sub>42</sub>H<sub>48</sub>N<sub>13</sub>O<sub>9</sub>  $[M+H]^+$  878.3698, found 878.3668.

#### 2.4.12 ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-Im-CO<sub>2</sub>H (**17**)

A solution of ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-Im-CO<sub>2</sub>Et **16** (2.0 g, 2.28 mmol) dissolved in 1,4-dioxane (2 mL), MeOH (6 mL), and aqueous KOH (1 N, 9.1 mL, 9.1 mmol) was stirred at 42 °C for 2 h. The solution was then acidified with aqueous HCl (1 N, ~9.1 mL, ~9.1 mmol) to a pH = 4.5, yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (10 mL) and collected by centrifugation (repeated 1x). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-Im-CO<sub>2</sub>H **17** as a tan solid (1.78 g, 92%). <sup>1</sup>H NMR: δ 10.50 (s, 1H), 10.47 (s, 1H), 9.96 (s, 1H), 9.92 (s, 1H), 7.98 (m, 1H), 7.48 (s, 1H), 7.40 (s, 1H), 7.30-7.27 (m, 6H), 7.24 (m, 1H), 7.19-7.14 (m, 3H), 7.05 (m, 2H), 6.90 (m, 1H), 4.99 (m, 2H), 4.10 (m, 1H), 3.99 (s, 3H), 3.88 (s, 3H), 3.845 (s, 3H), 3.838 (s, 3H), 3.77 (s, 3H), 3.30 (m, 2H), 2.49 (m, 2H, obstructed by NMR solvent); <sup>13</sup>C NMR: δ 167.7, 161.5, 160.0, 158.5, 155.8, 155.6, 138.6, 137.1, 131.6, 128.3, 127.7, 127.6, 126.7, 126.4, 123.0, 122.8, 122.7, 122.23, 122.16, 121.4, 118.7, 118.5, 118.0, 114.6, 105.0, 104.7, 104.4, 65.2, 48.6, 42.2, 38.2, 36.13, 36.10, 36.0, 35.4, 35.2; HRMS (FAB<sup>+</sup>) calc'd for C<sub>40</sub>H<sub>44</sub>N<sub>13</sub>O<sub>9</sub> [M+H]<sup>+</sup> 850.3385, found 850.3383.

#### 2.4.13 ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-ImPyPyPy-CO<sub>2</sub>Me (**2**)

A solution of ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-Im-CO<sub>2</sub>H **17** (1.5 g, 1.77 mmol) and PyBOP (546 mg, 1.85 mmol) in DMF (8.8 mL) and DIEA (922 μL, 5.3 mmol) was stirred at 23 °C for 10 min. The solution was then treated with solid (powdered) HCl•H<sub>2</sub>N-PyPyPy-CO<sub>2</sub>Me **4** (806 mg, 1.85 mmol) and stirred at 23 °C for 10 h. The solution was then added to distilled H<sub>2</sub>O (35 mL) preacidified with aqueous HCl (1 N, 5 mL, 5 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (40 mL) and collected by centrifugation (repeated 3x). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-ImPyPyPy-CO<sub>2</sub>Me **2** as a tan solid (2.09 g, 96%). <sup>1</sup>H NMR: δ 10.66 (s, 1H), 10.21 (s, 1H), 10.00 (s, 1H), 9.98 (s, 1H), 9.96 (s, 1H), 9.93 (s, 1H), 9.92 (s, 1H), 8.00 (m, 1H), 7.48 (s, 1H), 7.46 (d, *J* = 2.0 Hz, 1H), 7.45 (s, 1H), 7.31-7.27

(m, 7H), 7.23 (m, 2H), 7.20-7.14 (m, 5H), 7.06 (m, 2H), 6.92 (m, 1H), 6.90 (d,  $J = 2.0$  Hz, 1H), 5.00 (m, 2H), 4.11 (m, 1H), 4.00 (s, 3H), 3.95 (s, 3H), 3.854 (s, 3H), 3.850 (s, 3H), 3.842 (s, 3H), 3.837 (s, 3H), 3.828 (s, 3H), 3.78 (s, 3H), 3.73 (s, 3H), 3.32 (m, 2H), 2.53 (m, 2H);  $^{13}\text{C}$  NMR:  $\delta$  167.9, 161.6, 160.8, 158.5, 158.42, 158.38, 155.8, 155.6, 137.6, 137.1, 136.0, 134.0, 128.3, 127.7, 127.6, 126.4, 123.3, 123.1, 123.0, 122.80, 122.77, 122.5, 122.26, 122.24, 122.1, 121.2, 120.9, 120.8, 118.9, 118.70, 118.64, 118.5, 118.0, 114.1, 108.4, 104.9, 104.79, 104.76, 104.5, 65.2, 51.0, 48.8, 42.2, 38.4, 36.25, 36.21, 36.19, 36.13, 36.10, 36.0, 35.8, 35.0; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{59}\text{H}_{64}\text{N}_{19}\text{O}_{12}$  [M+H]<sup>+</sup> 1230.498, found 1230.504.

#### 2.4.14 *ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-ImPyPyPy-CO<sub>2</sub>H* (**22**)

A solution of polyamide **2** (500 mg, 0.406 mmol) dissolved in 1,4-dioxane (8 mL) and aqueous NaOH (1 N, 8.0 mL, 8.0 mmol) was stirred at 23 °C for 11 h. The solution was then cooled to 0 °C in an ice bath and the pH adjusted to pH = 4.0 with aqueous HCl (1 N, ~8 mL, 8 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (40 mL) and collected by centrifugation (repeated 2x). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield *ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-ImPyPyPy-CO<sub>2</sub>H* **22** as a tan solid (442 mg, 89%).  $^1\text{H}$  NMR:  $\delta$  10.52 (s, 1H), 10.22 (s, 1H), 10.00 (s, 1H), 9.97 (s, 1H), 9.95 (s, 1H), 9.93 (s, 1H), 9.90 (s, 1H), 8.00 (m, 1H), 7.45 (s, 1H), 7.42 (m, 2H), 7.30-7.27 (m, 7H), 7.24 (m, 2H), 7.19-7.14 (m, 4H), 7.09 (s, 1H), 7.06 (m, 2H), 6.92 (m, 1H), 6.84 (d,  $J = 2.0$  Hz, 1H), 5.01 (m, 2H), 4.11 (m, 1H), 3.99 (s, 3H), 3.95 (s, 3H), 3.851 (s, 3H), 3.848 (s, 3H), 3.843 (s, 3H), 3.838 (s, 3H), 3.81 (s, 3H), 3.79 (s, 3H), 3.32 (m, 2H), 2.53 (m, 2H);  $^{13}\text{C}$  NMR:  $\delta$  167.9, 162.0, 161.6, 158.48, 158.47, 158.44, 158.40, 155.8, 155.6, 138.6, 137.1, 136.0, 134.0, 128.3, 127.7, 127.6, 126.4, 123.06, 123.05, 122.8, 122.74, 122.70, 122.6, 122.24, 122.19, 122.15, 121.4, 121.2, 120.3, 119.5, 118.7, 118.55, 118.47, 118.0, 114.1, 108.4, 104.9, 104.86, 104.80, 104.75, 104.5, 65.2, 48.7, 42.2, 38.3, 36.2, 36.13, 36.10, 36.07, 35.97, 35.2, 34.9; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{58}\text{H}_{62}\text{N}_{19}\text{O}_{12}$  [M+H]<sup>+</sup> 1216.483, found 1216.487.

#### 2.4.15 *BocHN-(+)-BnOIPA* (**20**)

A solution of acid **19**<sup>15</sup> (211 mg, 0.824 mmol) and PyBOP (624 mg, 1.2 mmol) in DMF (3 mL) and DIEA (211  $\mu\text{L}$ , 1.2 mmol) was stirred at 23 °C for 10 min. Protected triamine **18**<sup>14</sup> (373 mg, 1.52

mmol) was then added to the solution and stirring was continued at 23 °C for 3 h. The solution was then added to distilled H<sub>2</sub>O (15 mL) and a viscous oil was isolated following centrifugation (~4500 rpm) and decanting of the aqueous layer. The residual oil was again washed with distilled H<sub>2</sub>O (10 mL) and collected by centrifugation (repeated 3x). Drying of the residual oil *in vacuo* yielded BocHN-(+)-Bn<sup>0</sup>IPA **20** as a viscous amber oil (391 mg, 98%). <sup>1</sup>H NMR: δ 8.77 (t, *J* = 5.5 Hz, 1H), 8.43 (app t, *J* = 1.5 Hz, 1H), 8.14-8.09 (m, 2H), 7.63 (app t, *J* = 7.8 Hz, 1H), 7.48-7.34 (m, 5H), 6.85 (t, *J* = 4.9 Hz, 1H), 5.38 (s, 2H), 3.30 (m, 2H), 3.00 (m, 1H), 2.94 (m, 2H), 2.72-2.51 (m, 4H), 2.38 (br s, 3H), 1.75 (m, 2H), 1.59 (m, 2H), 1.34 (s, 9H); <sup>13</sup>C NMR: δ 165.3, 165.2, 155.6, 136.0, 135.0, 131.9, 131.7, 129.8, 129.0, 128.6, 128.2, 128.1, 127.9, 77.5, 66.4, 54.3, 54.0, 40.6, 37.7, 37.3, 28.2, 25.9, 25.5; HRMS (FAB<sup>+</sup>) calc'd for C<sub>27</sub>H<sub>38</sub>N<sub>3</sub>O<sub>5</sub> [M+H]<sup>+</sup> 484.2811, found 484.2793.

#### 2.4.16 ImPyPyPy-(*R*)<sup>β-CbzHN</sup>γ-ImPyPyPy-(+)-Bn<sup>0</sup>IPA (**23**)

A solution of ImPyPyPy-(*R*)<sup>β-CbzHN</sup>γ-ImPyPyPy-CO<sub>2</sub>H **22** (192 mg, 0.158 mmol) and PyBOP (99 mg, 0.189 mmol) in DMF (2 mL) and DIEA (220 μL, 1.3 mmol) was stirred at 23 °C for 30 min. In a separate vial, BocHN-(+)-Bn<sup>0</sup>IPA **20** (122 mg, 0.252 mmol) was deprotected by treating with a solution of CH<sub>2</sub>Cl<sub>2</sub>:CF<sub>3</sub>CO<sub>2</sub>H (1:1, 2 mL) for 20 minutes at 23 °C followed by concentration to dryness *in vacuo*. This residual material was then treated with the preactivated solution of ImPyPyPy-(*R*)<sup>β-CbzHN</sup>γ-ImPyPyPy-CO<sub>2</sub>H **22** and allowed to stir at 23 °C for 12 h. The solution was then added to distilled H<sub>2</sub>O (30 mL) preacidified with aqueous HCl (1 N, 2 mL, 2 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (40 mL) and collected by centrifugation (repeated 3x). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The residual brown solid was purified by SiO<sub>2</sub> chromatography with the mobile phase consisting of a step gradient of 49:1 CH<sub>2</sub>Cl<sub>2</sub>:MeOH to 44:5:1 CH<sub>2</sub>Cl<sub>2</sub>:MeOH:NH<sub>3</sub> to provide ImPyPyPy-(*R*)<sup>β-CbzHN</sup>γ-ImPyPyPy-(+)-Bn<sup>0</sup>IPA **23** as a tan solid (217 mg, 87%) after drying under high vacuum. <sup>1</sup>H NMR: δ 10.43 (s, 1H), 10.20 (s, 1H), 9.98 (s, 1H), 9.95 (s, 1H), 9.93 (s, 1H), 9.92 (s, 1H), 9.87 (s, 1H), 8.73 (t, *J* = 5.5 Hz, 1H), 8.41 (app t, *J* = 1.6 Hz, 1H), 8.11-8.07 (m, 2H), 8.03-7.98 (m, 2H), 7.60 (app t, *J* = 7.7 Hz, 1H), 7.48-7.32 (m, 7H), 7.30-7.23 (m, 9H), 7.19-7.14 (m, 5H), 7.06 (d, *J* = 1.7 Hz, 1H), 7.05 (d, *J* = 1.7 Hz, 1H), 7.03 (d, *J* = 1.0 Hz, 1H), 6.92 (d, *J* = 1.2 Hz, 1H), 6.84 (d, *J* = 1.7 Hz, 1H), 5.37 (s, 2H), 5.00 (m, 2H), 4.11 (m, 1H), 3.99 (s, 3H), 3.95 (s, 3H), 3.85 (s, 3H), 3.84 (m, 6H), 3.83 (s, 3H), 3.79 (s, 3H), 3.77 (s, 3H), 3.30 (m, 4H), 3.18 (m, 2H), 2.53 (m, 2H), 2.35 (m, 4H), 2.16 (br s, 3H), 1.71-1.60 (m, 4H); <sup>13</sup>C NMR: δ 167.9, 165.20, 165.16, 161.6, 161.2, 158.50,

158.46, 158.42, 156.1, 155.8, 155.6, 138.8, 137.1, 136.0, 135.2, 134.0, 131.9, 131.6, 129.7, 129.0, 128.5, 128.3, 128.2, 128.1, 128.0, 127.8, 127.7, 127.6, 127.00, 126.99, 126.3, 123.1, 123.0, 122.80, 122.77, 122.75, 122.26, 122.18, 122.14, 122.11, 121.4, 121.2, 118.7, 118.5, 118.0, 117.7, 114.1, 105.0, 104.9, 104.8, 104.5, 104.1, 66.4, 65.2, 55.1, 55.0, 48.8, 42.2, 41.6, 38.3, 37.9, 37.0, 36.2, 36.10, 36.09, 36.07, 35.96, 35.89, 35.1, 34.9, 26.9, 26.6; HRMS (ESI<sup>+</sup>) calc'd for C<sub>80</sub>H<sub>89</sub>N<sub>22</sub>O<sub>14</sub> [M+H]<sup>+</sup> 1581.6929, found 1581.6992.

#### 2.4.17 ImPyPyPy-(R)<sup>β-H<sub>2</sub>N</sup>γ-ImPyPyPy-(+)-IPA (**1**)

A solution of ImPyPyPy-(R)<sup>β-CbzHN</sup>γ-ImPyPyPy-(+)-Bn<sup>O</sup>IPA **23** (25 mg, 0.016 mmol) and Pd/C (10 wt % dry, 10 mg) in DMF (2 mL) was stirred at 23 °C for 48 h under H<sub>2</sub> (~1 atm, balloon). [Note: The protecting groups are cleaved at different rates, and early reaction aliquots reveal a mono-protected compound.] The reaction was then filtered through a Sep-Pak cartridge (5 g of C-18 sorbent), and the Sep-Pak was washed with DMF (4 mL), aqueous MeCN (50%, 20 mL), MeCN (250 mL), and MeOH (250 mL). The filtrate was then concentrated *in vacuo*, purified by reverse-phase HPLC, and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield ImPyPyPy-(R)<sup>β-H<sub>2</sub>N</sup>γ-ImPyPyPy-(+)-IPA **1** as a light tan solid (18.9 mg, 81%). <sup>1</sup>H NMR: δ 10.61 (s, 1H), 10.46 (s, 1H), 9.95 (s, 2H), 9.94 (s, 1H), 9.92 (s, 1H), 9.89 (s, 1H), 9.46 (br s, 1H), 8.82 (t, *J* = 5.8 Hz, 1H), 8.42 (app t, *J* = 1.4 Hz, 1H), 8.20 (m, 1H), 8.15 (m, 1H), 8.08 (d, *J* = 1.7 Hz, 1H), 8.07 (d, *J* = 1.7 Hz, 1H), 7.97 (m, 3H), 7.60 (app t, *J* = 7.6 Hz, 1H), 7.45 (s, 1H), 7.40 (d, *J* = 0.7 Hz, 1H), 7.28 (d, *J* = 2.0 Hz, 1H), 7.26 (d, *J* = 1.7 Hz, 1H), 7.22 (d, *J* = 2.0 Hz, 1H), 7.21 (d, *J* = 1.7 Hz, 1H), 7.18 (d, *J* = 1.7 Hz, 1H), 7.16 (m, 2H), 7.15 (d, *J* = 1.7 Hz, 1H), 7.08 (m, 2H), 7.05 (d, *J* = 1.0 Hz, 1H), 7.03 (d, *J* = 1.7 Hz, 1H), 6.94 (d, *J* = 1.8 Hz, 1H), 3.99 (s, 3H), 3.96 (s, 3H), 3.85 (s, 3H), 3.844 (s, 3H), 3.842 (s, 3H), 3.83 (s, 3H), 3.82 (s, 3H), 3.79 (s, 3H), 3.69 (m, 1H), 3.44 (m, 2H), 3.35 (m, 2H), 3.24 (m, 2H), 3.18-3.12 (m, 2H), 3.11-3.04 (m, 2H), 2.76 (m, 3H), 2.50 (m, 2H, obstructed by NMR solvent), 1.95-1.91 (m, 2H), 1.90-1.83 (m, 2H); <sup>13</sup>C NMR: δ 166.84, 166.82, 165.7, 162.1, 161.6, 158.52, 158.49, 158.40, 158.1, 157.8, 156.0, 155.7, 138.7, 135.6, 134.6, 134.2, 131.9, 131.5, 131.0, 128.7, 128.0, 126.9, 126.3, 123.1, 123.0, 122.8, 122.7, 122.5, 122.4, 122.24, 122.16, 122.10, 121.4, 121.1, 118.7, 118.5, 118.3, 118.0, 115.7, 105.0, 104.9, 104.8, 104.54, 104.52, 59.2, 53.3, 48.2, 36.5, 36.2, 36.08, 36.07, 35.99, 35.6, 35.38, 35.35, 35.1, 35.0, 24.3, 24.0; HRMS (ESI<sup>+</sup>) calc'd for C<sub>65</sub>H<sub>77</sub>N<sub>22</sub>O<sub>12</sub> [M+H]<sup>+</sup> 1357.6086, found 1357.6091.

#### 2.4.18 Calculation of Molar Extinction Coefficients

Solid **1** (0.51 mg, 0.35  $\mu$ moles) was weighed into a tared vial on a microbalance (Sartorius Micro M4). A molecular weight of 1471.46 was utilized for **1**, which assumes **1** exists as the  $\text{CF}_3\text{CO}_2\text{H}$  salt. The material was dissolved in distilled and deionized  $\text{H}_2\text{O}$  (1000  $\mu\text{L}$ ), yielding a 0.35  $\mu\text{M}$  stock solution. Fifty-fold dilutions of this stock solution were made into the appropriate solvent systems (Figure 2.4). Data were collected at 23  $^\circ\text{C}$  in a 1 cm quartz cuvette, and molar extinction coefficients are based on the  $\lambda_{\text{max}}$  for each solvent system (range: 313–317 nm). The instrument was blanked on each discrete solvent system prior to data collection. Duplicate analysis of each sample yielded similar results (data not shown). A second polyamide stock solution was generated by dissolving solid **1** in 50% MeCN in aqueous  $\text{CF}_3\text{CO}_2\text{H}$  (0.1% v/v), followed by 50-fold dilution into the individual solvent systems, and data collection yielded similar results to those shown in Figure 2.4 (data not shown). DEPC-treated  $\text{H}_2\text{O}$  (RNase- and DNase-free) was purchased from USB Corp. EtOH was absolute grade from Pharmaco-AAPER, and DMSO was molecular biology grade from Sigma-Aldrich. DMF was anhydrous synthesis grade from Sigma-Aldrich, and acetonitrile was HPLC grade from Fisher.

## 2.5 Notes and References

1. Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215–2235.
2. Dervan, P. B., and Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–299.
3. (a) Trauger, J. W., Baird, E. E., and Dervan, P. B. *Nature* **1996**, *382*, 559–561. (b) White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. *Nature* **1998**, *391*, 468–470. (c) Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104–109. (d) Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. *Science* **1998**, *282*, 111–115.
4. Hsu, C. F., Phillips, J. W., Trauger, J. W., Farkas, M. E., Belitsky, J. M., Heckel, A., Olenyuk, B. Z., Puckett, J. W., Wang, C. C. C., and Dervan, P. B. *Tetrahedron* **2007**, *63*, 6146–6151.
5. (a) Belitsky, J. M., Leslie, S. J., Arora, P. S., Beerman, T. A., and Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 3313–3318. (b) Crowley, K. S., Phillion, D. P., Woodard, S. S., Scheitzer, B. A., Singh, M., Shabany, H., Burnette, B., Hippenmeyer, P., Heitmeier, M., and Bashkin, J. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1565–1570. (c) Best, T. P., Edelson, B. S., Nickols, N. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12063–12068. (d) Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., and Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802–2818. (e) Xiao, X., Yu, P., Lim, H. S., Sikder, D., and Kodadek, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865–2868. (f) Nickols, N. G., Jacobs, C. S., Farkas, M. E., and Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 363–370. (g) Dose, C., Farkas, M. E., Chenoweth, D. M., and Dervan, P. B. *J. Am. Chem. Soc.* **2008**, *130*, 6859–6866. (h) Hsu, C. F., and Dervan, P. B. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5851–5855.
6. (a) Gottesfeld, J. M., Melander, C., Suto, R. K., Raviol, H., Luger, K., and Dervan, P. B. *J. Mol. Biol.* **2001**, *309*, 615–629. (b) Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. *J. Mol. Biol.* **2003**, *326*, 371–380. (c) Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6864–6869. (d) Dudouet, B., Burnett, R., Dickinson, L. A., Wood, M. R., Melander, C., Belitsky, J. M., Edelson, B., Wurtz, N., Briehn, C., Dervan, P. B., and Gottesfeld, J. M. *Chem. Biol.* **2003**, *10*, 859–867.
7. (a) Olenyuk, B. Z., Zhang, G. J., Klco, J. M., Nickols, N. G., Kaelin, Jr., W. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16768–16773. (b) Kageyama, Y., Sugiyama, H., Ayame, H., Iwai, A., Fujii, Y., Huang, L. E., Kizaka-Kondoh, S., Hiraoka, M., and Kihara, K. *Acta Oncol.* **2006**, *45*, 317–324. (c) Nickols, N. G., Jacobs, C. S., Farkas, M. E., and Dervan, P. B. *ACS Chem. Biol.* **2007**, *2*, 561–571.
8. Nickols, N. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10418–10423.
9. (a) Matsuda, H., Fukuda, N., Ueno, T., Tahira, Y., Ayame, H., Zhang, W., Bando, T., Sugiyama, H., Saito, S., Matsumoto, K., and others, O. *J. Am. Soc. Nephrol.* **2006**, *17*, 422–432. (b) Yao, E. H., Fukuda, N., Ueno, T., Matsuda, H., Matsumoto, K., Nagase, H., Matsumoto, Y., Takasaka, A., Serie, K., Sugiyama, H., and Sawamura, T. *Hypertension* **2008**, *52*, 86–92.
10. (a) Baird, E. E., and Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141–6146. (b) Belitsky, J. M., Nguyen, D. H., Wurtz, N. R., and Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 2767–2774. (c) Wurtz, N. R., Turner, J. M., Baird, E. E., and Dervan, P. B. *Org. Lett.* **2001**, *3*, 1201–1203. (d) Krutzik, P. O., and Chamberlin, A. R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2129–2132. (e) Krutzik, P. O., and Chamberlin, A. R. *Methods Mol. Biol.* **2002**, *201*, 77–92. (f) Ayame, H., Saito, T., Bando, T., Fukuda, N., and Sugiyama, H. *Nucleic Acids Res. Suppl.* **2003**, 67–68. (g) Moore, M. J. B.;

Cuenca, F.; Searcey, M.; Neidle, S. *Org. Biomol. Chem.* **2006**, *4*, 3479–3488.

11. (a) Mrksich, M., Parks, M. E., and Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983–7988. (b) Xiao, J., Yuan, G., Huang, W., Chan, A. S., and Lee, K. L. *J. Org. Chem.* **2000**, *65*, 5506–5513. (c) Harris, D., Stewart, M., Sielaff, A., Mulder, K., Brown, T., Mackay, H., and Lee, M. *Heterocycl. Commun.* **2007**, *13*, 17–23. (d) Xiao, J.-H., Huang, W.-Q., Tang F.-L., Yuan G., Chan A. S. C., and Lee K.-L. *Chin. J. Chem.* **2000**, *18*, 603–607. (e) Boger, D. L.; Fink, B. E.; Hedrick, M. P. *J. Am. Chem. Soc.* **2000**, *122*, 6382–6394. (f) Mamidyala, S. K.; Firestine, S. M. *Tetrahedron Lett.* **2006**, *47*, 7431–7434.

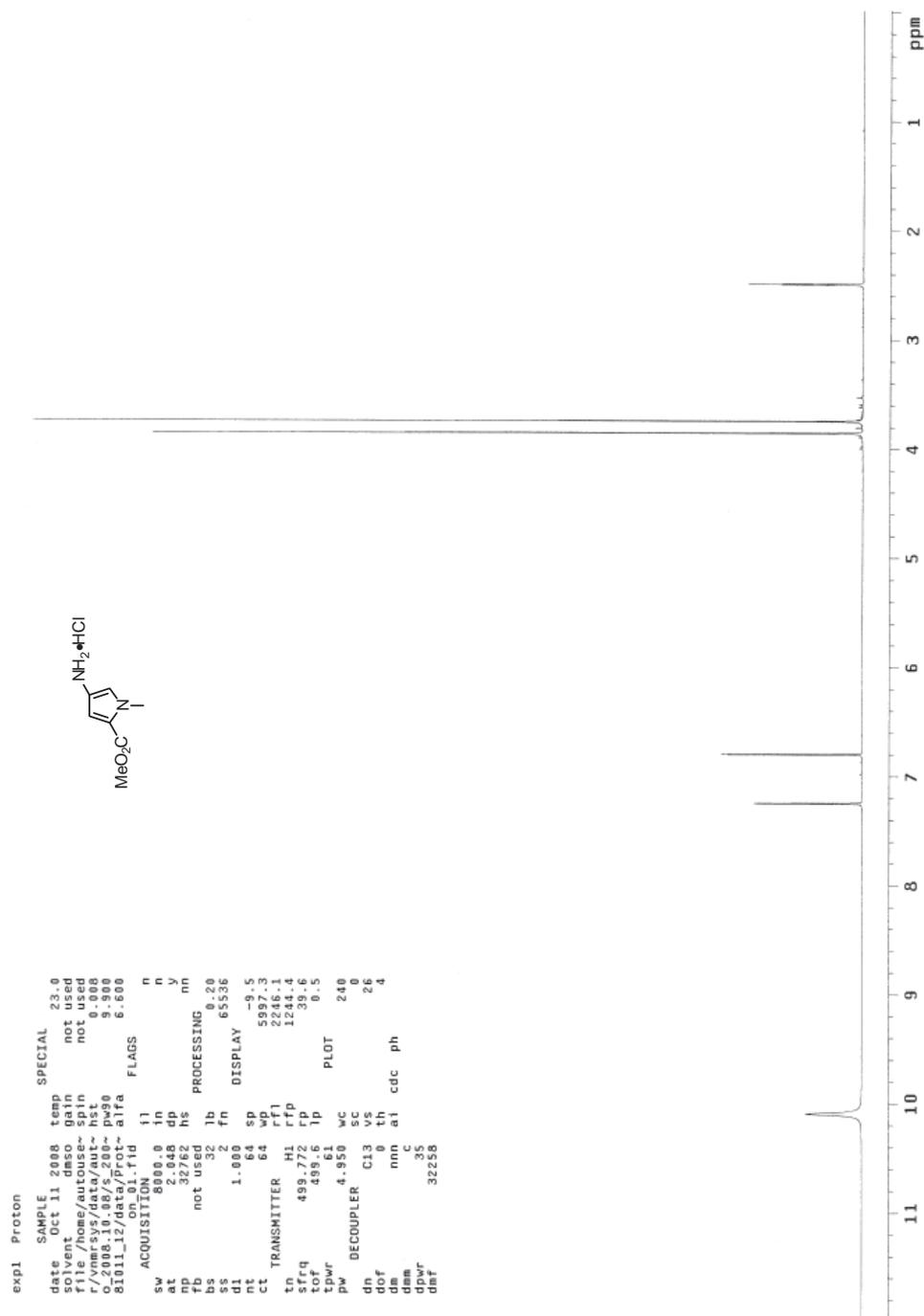
12. Sources for commercially available building blocks **7–10** shown in Figure 2.2; **7** (Fluorochem, Cat # 019570), **8** (Fluka, Cat # 17974), **9** (Bachem, Cat # F-3480), **10** (Bachem, Cat # F-3485).

13. Jaramillo, D., Liu, Q., Aldrich-Wright, J., and Tor, Y. *J. Org. Chem.* **2004**, *69*, 8151–8153.

14. Dervan, P. B., Sluka, J. P. Design of sequence specific DNA binding molecules: bis(distamycin) phenoxazine. In “*New Synthetic Methodology and Functionally Interesting Compounds*”; Yoshida, Z., Ed.; Studies in Organic Chemistry 25; Elsevier: Amsterdam, 1986; pp 307–322.

15. Adlington, R. M., Baldwin, J. E., Becker, G. W., Chen, B., Cheng, L., Cooper, S. L., Hermann, R. B., Howe, T. J., McCoull, W., McNulty, A. M., Neubauer, B. L., and Pritchard, G. J. *J. Med. Chem.* **2001**, *44*, 1491–1508.

## 2.6 Spectra and Supplemental Information

 $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra for synthesized compounds.Figure 2.5  $^1\text{H}$  NMR of  $\text{HCl}\cdot\text{H}_2\text{N}\text{-Py-CO}_2\text{Me}$  (10)

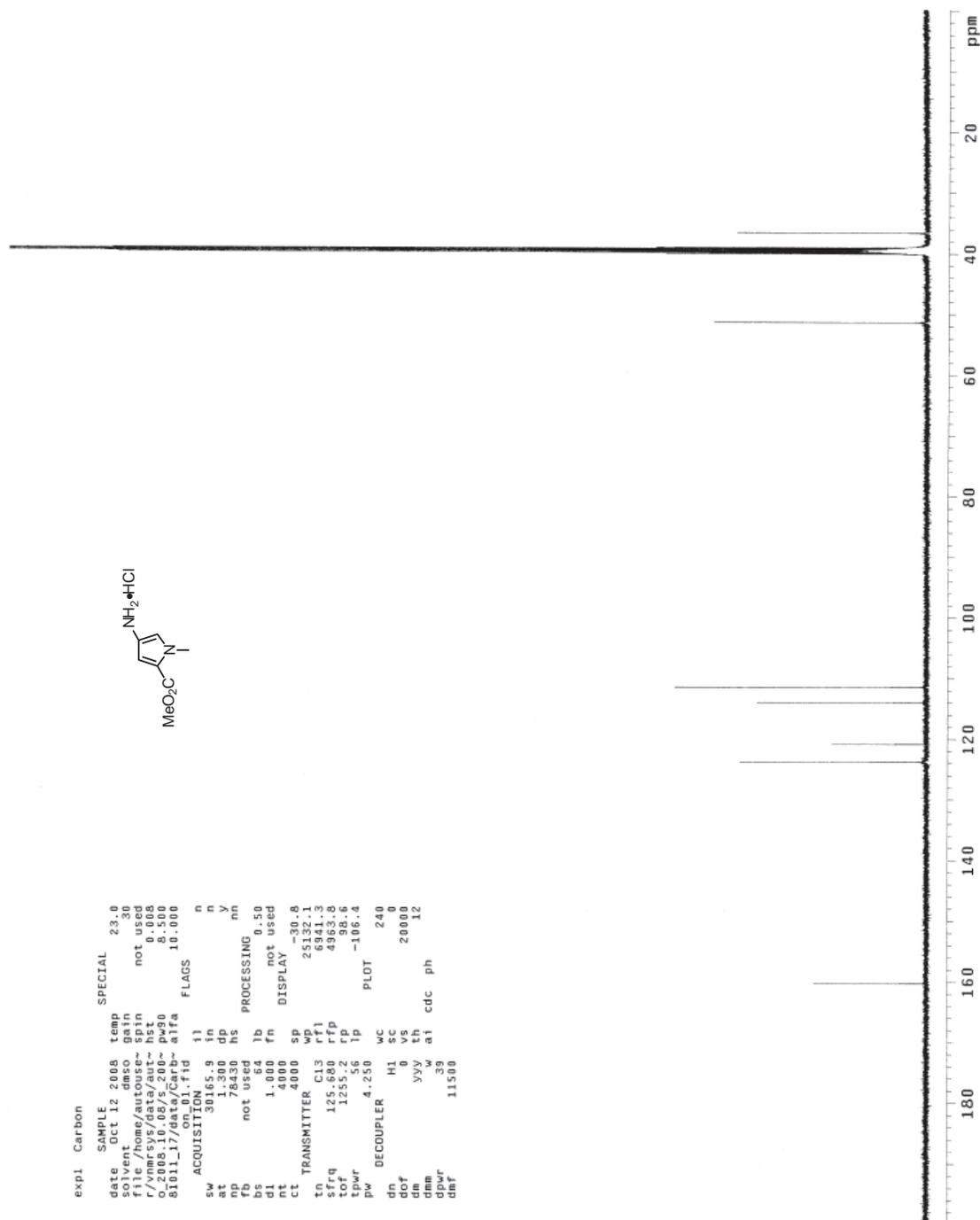


Figure 2.6  $^{13}\text{C}$  NMR of  $\text{HCl}\cdot\text{H}_2\text{N-Py-CO}_2\text{Me}$  (10)

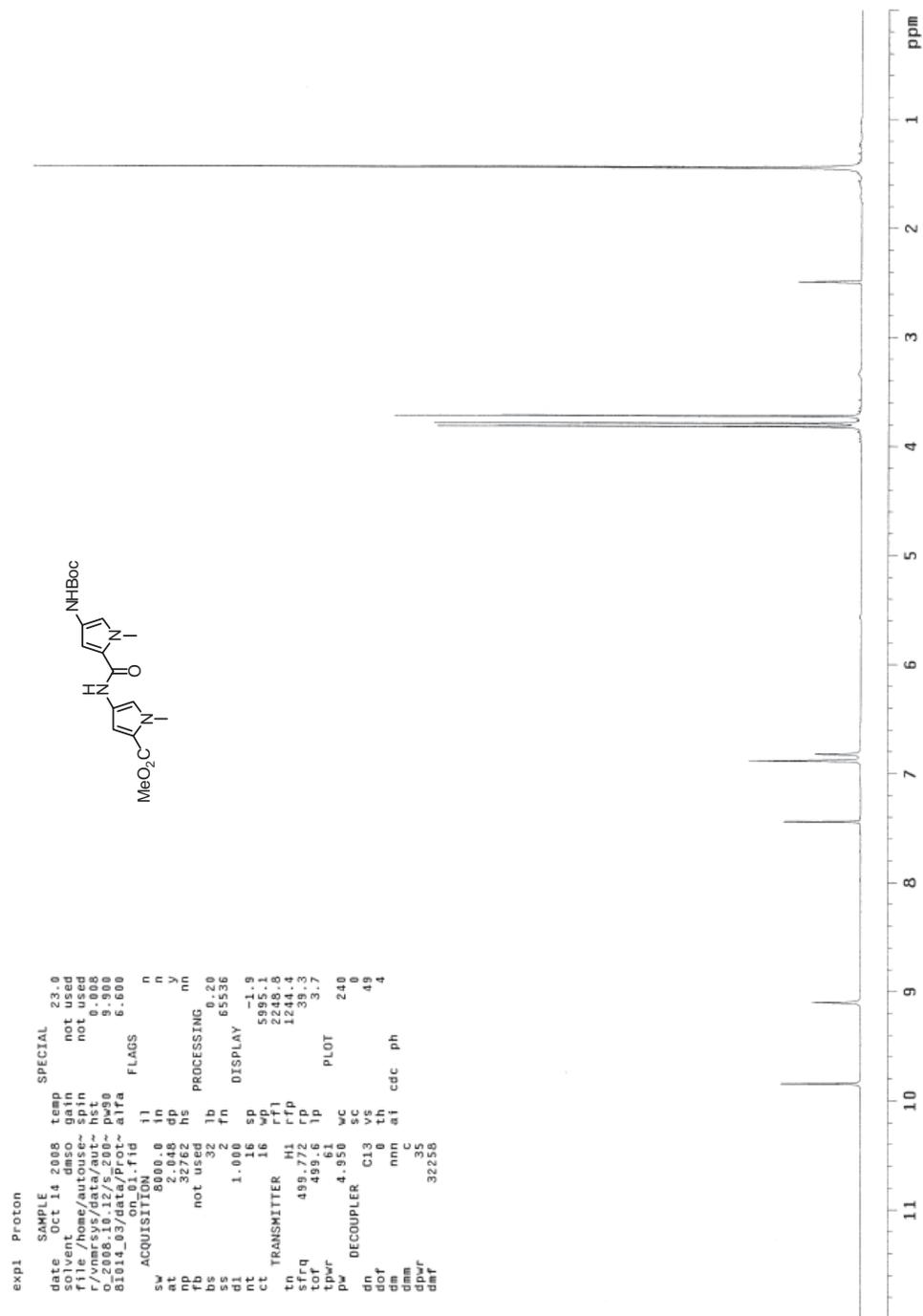
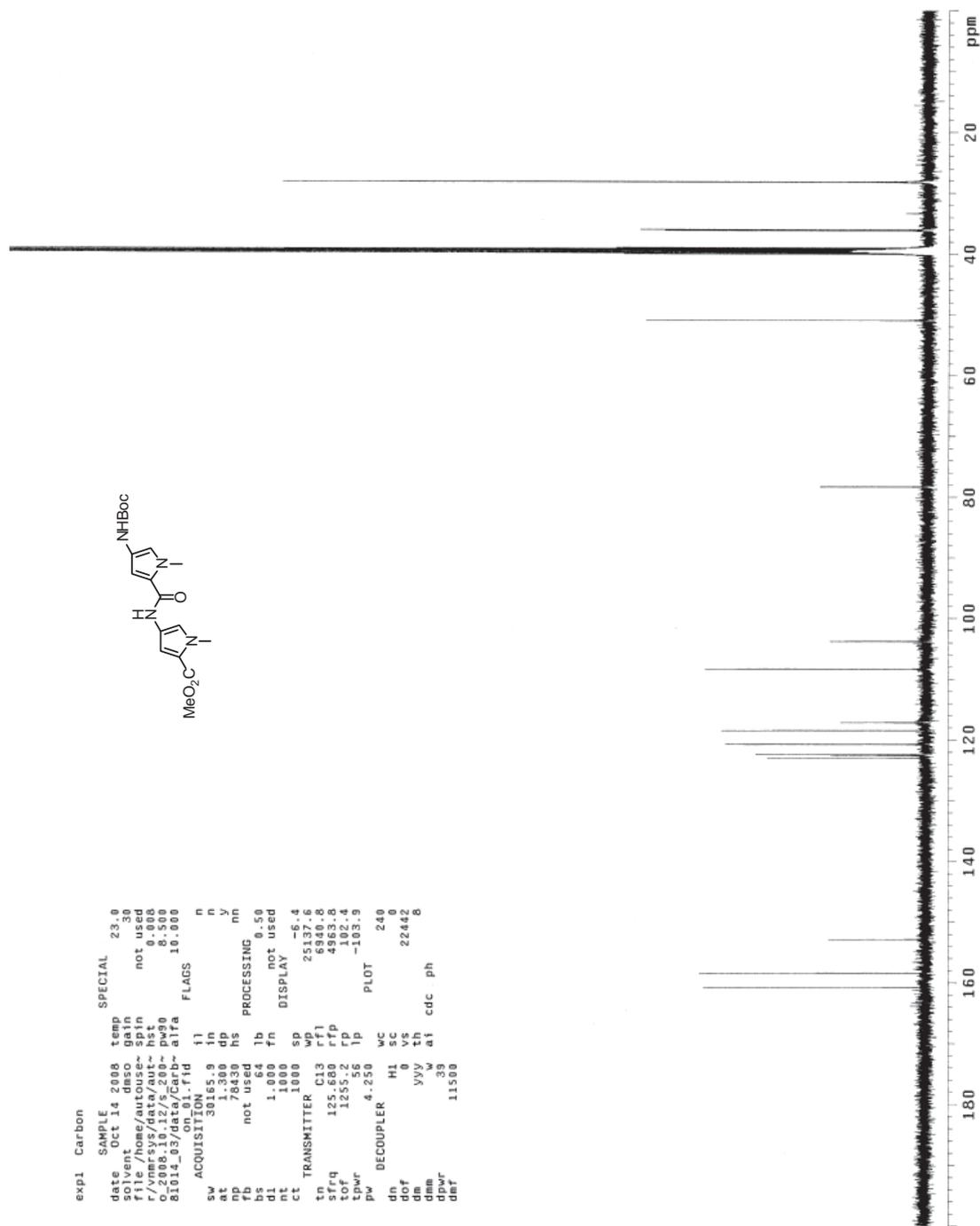


Figure 2.7 <sup>1</sup>H NMR of BocHN-PyPy-CO<sub>2</sub>Me (11)



**Figure 2.8**  $^{13}\text{C}$  NMR of BocHN-PyPy- $\text{CO}_2\text{Me}$  (11)

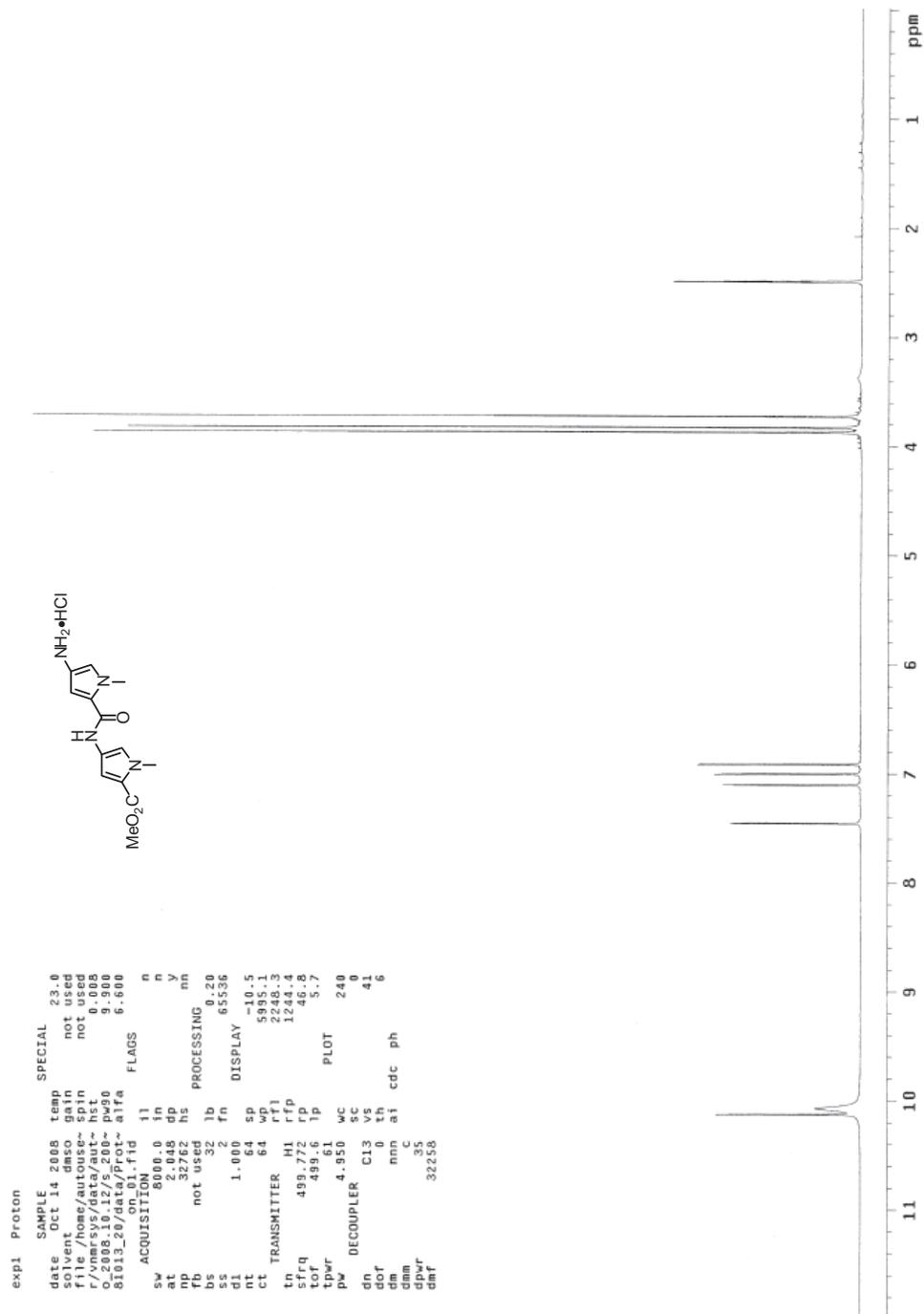


Figure 2.9 <sup>1</sup>H NMR of HCl•H<sub>2</sub>N-PyPy-CO<sub>2</sub>Me (12)

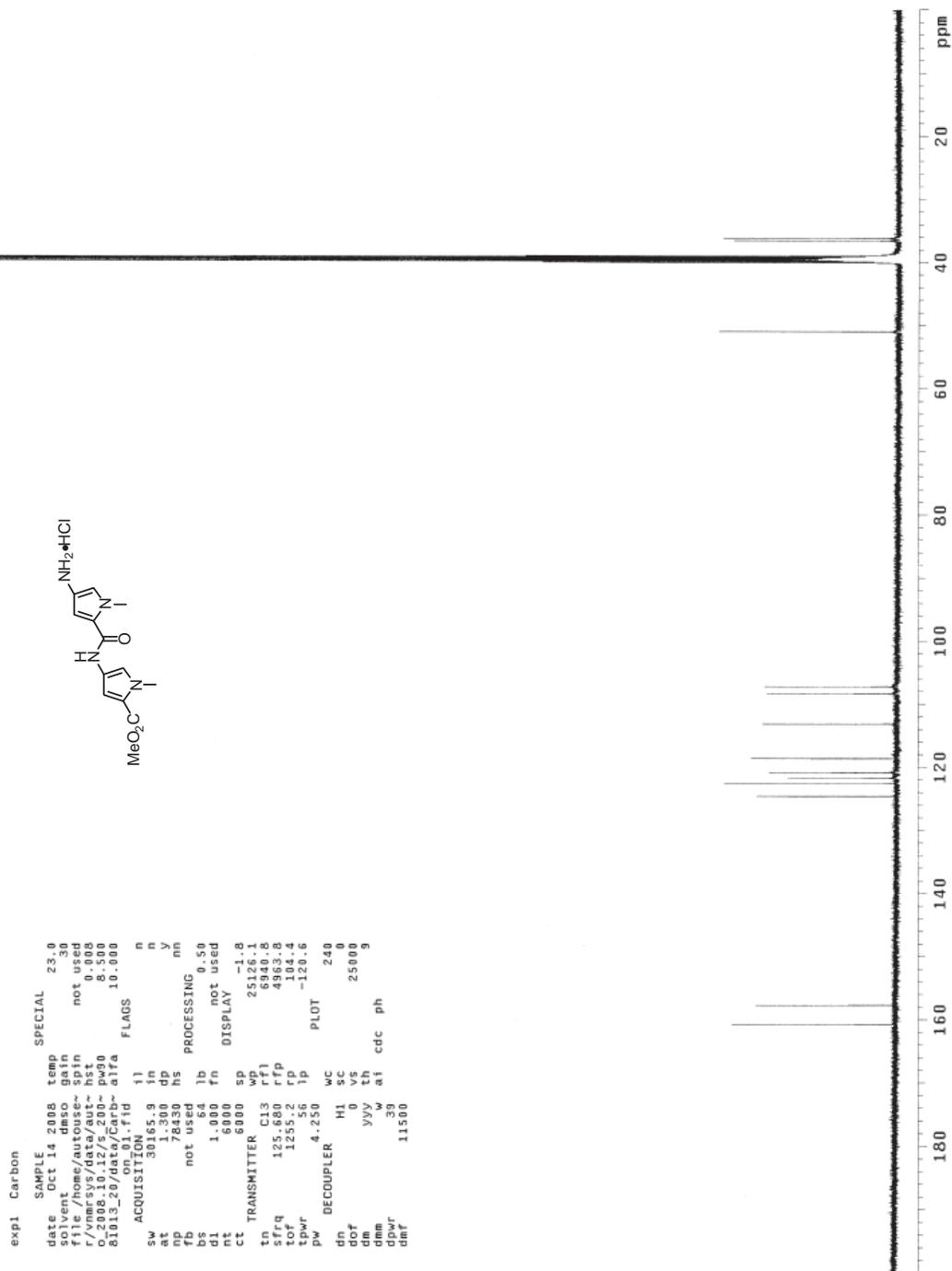


Figure 2.10 <sup>13</sup>C NMR of HCl•H<sub>2</sub>N-PyPy-CO<sub>2</sub>Me (12)

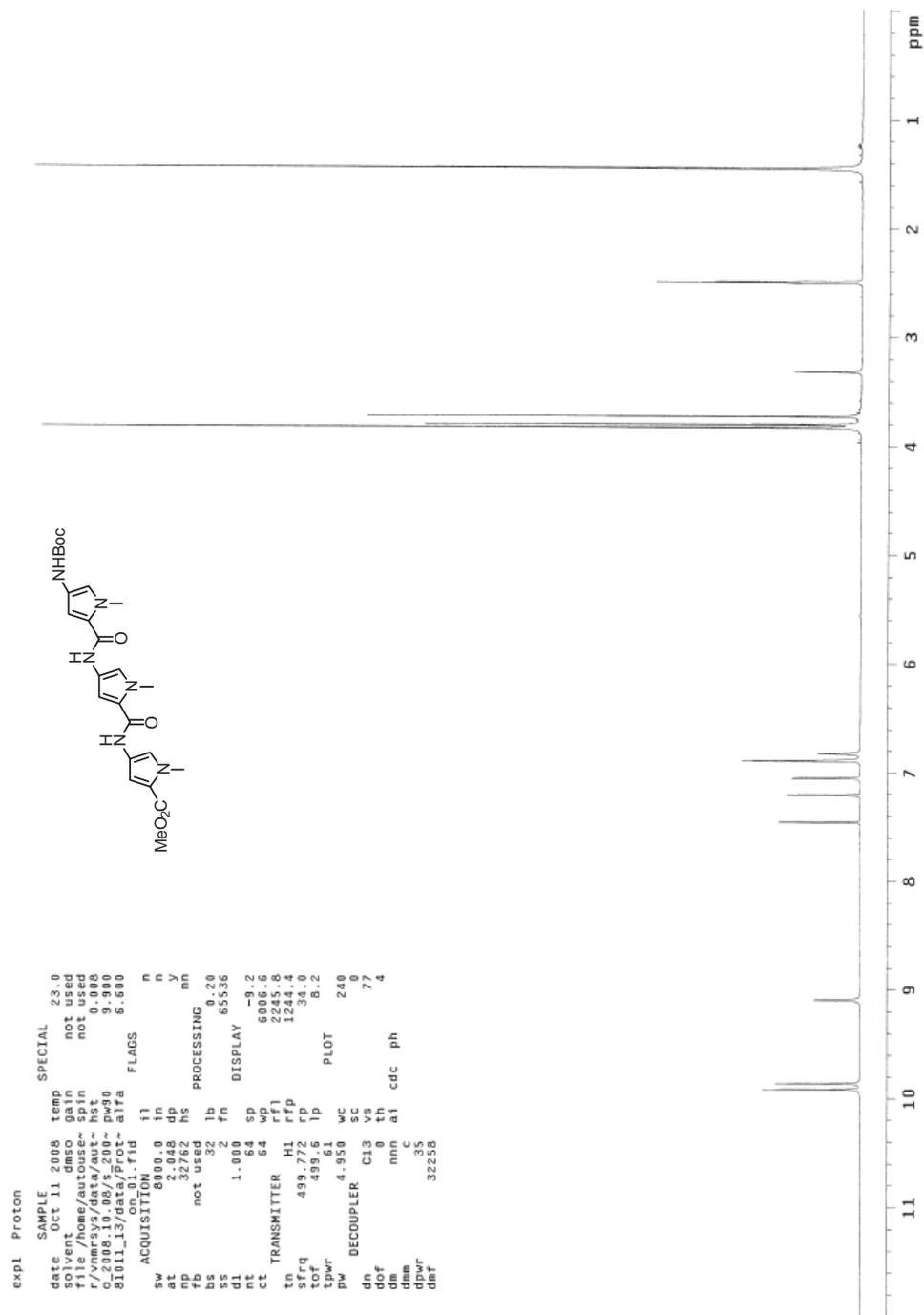


Figure 2.11 <sup>1</sup>H NMR of BocHN-PyPyPy-CO<sub>2</sub>Me (13)

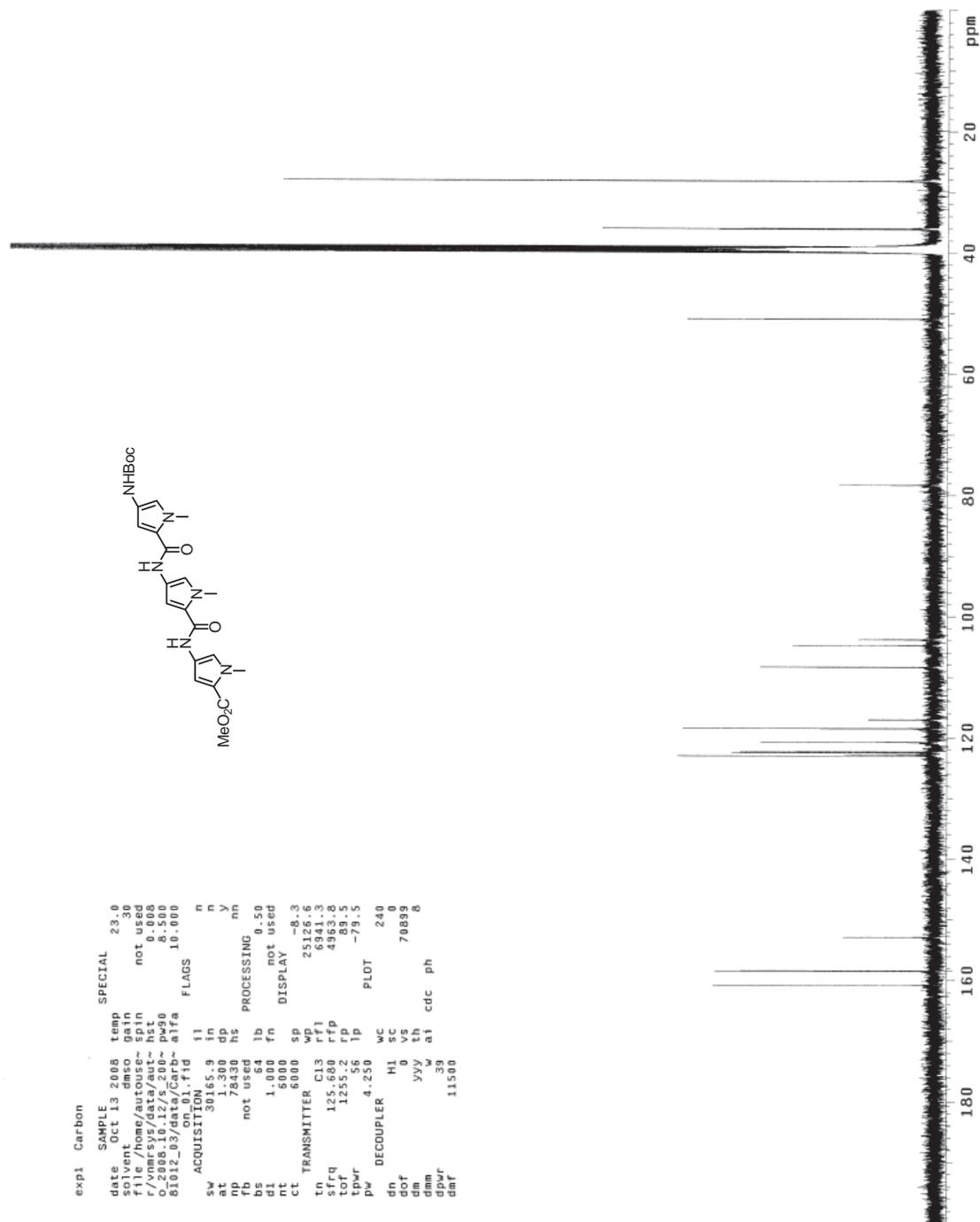


Figure 2.12  $^{13}\text{C}$  NMR of BocHN-PyPyPy-CO<sub>2</sub>Me (13)

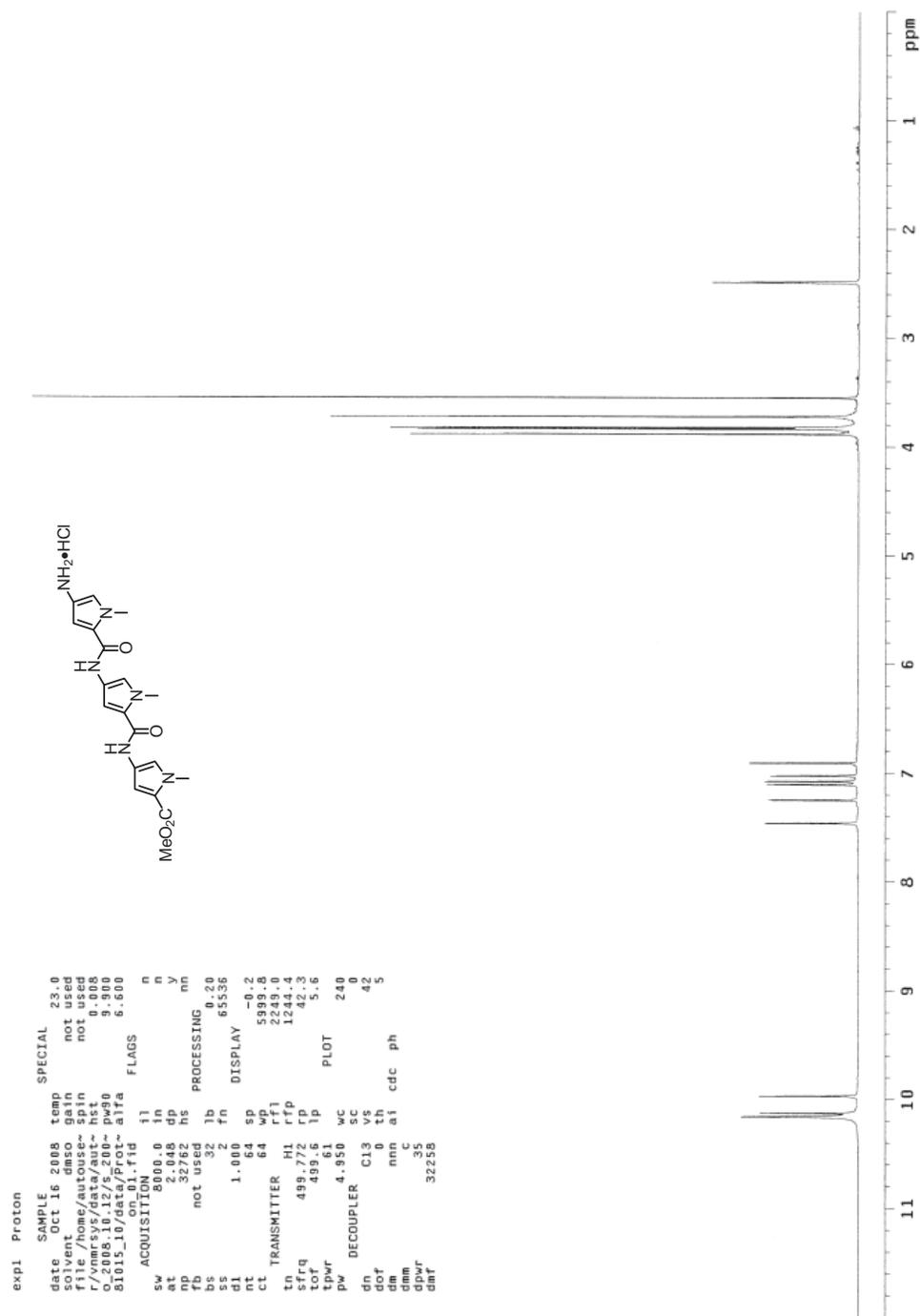


Figure 2.13  $^1\text{H}$  NMR of  $\text{HCl}\cdot\text{H}_2\text{N-PyPyPy-CO}_2\text{Me}$  (4)

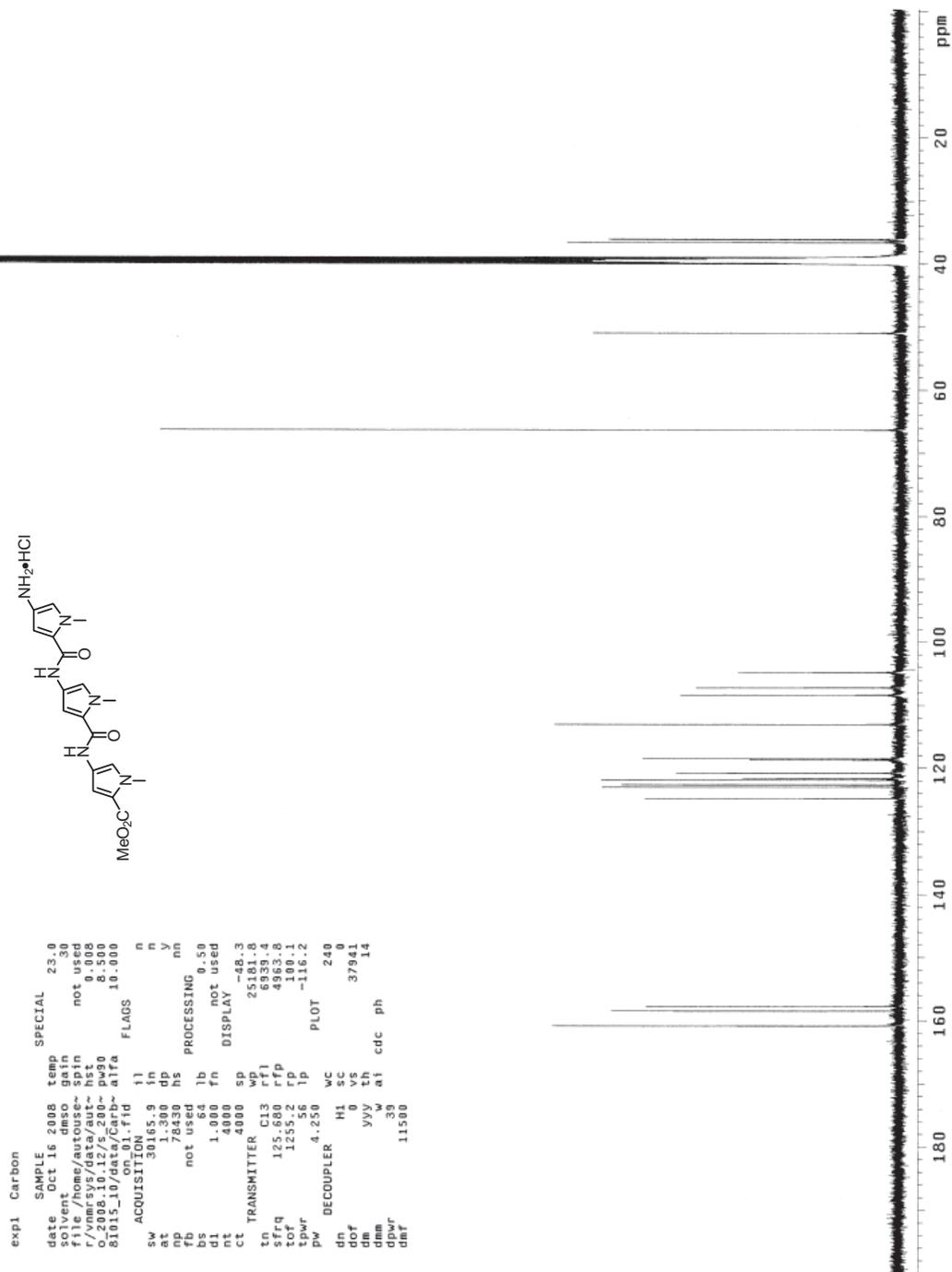


Figure 2.14  $^{13}\text{C}$  NMR of  $\text{HCl}\cdot\text{H}_2\text{N-PyPyPy-CO}_2\text{Me}$  (4)

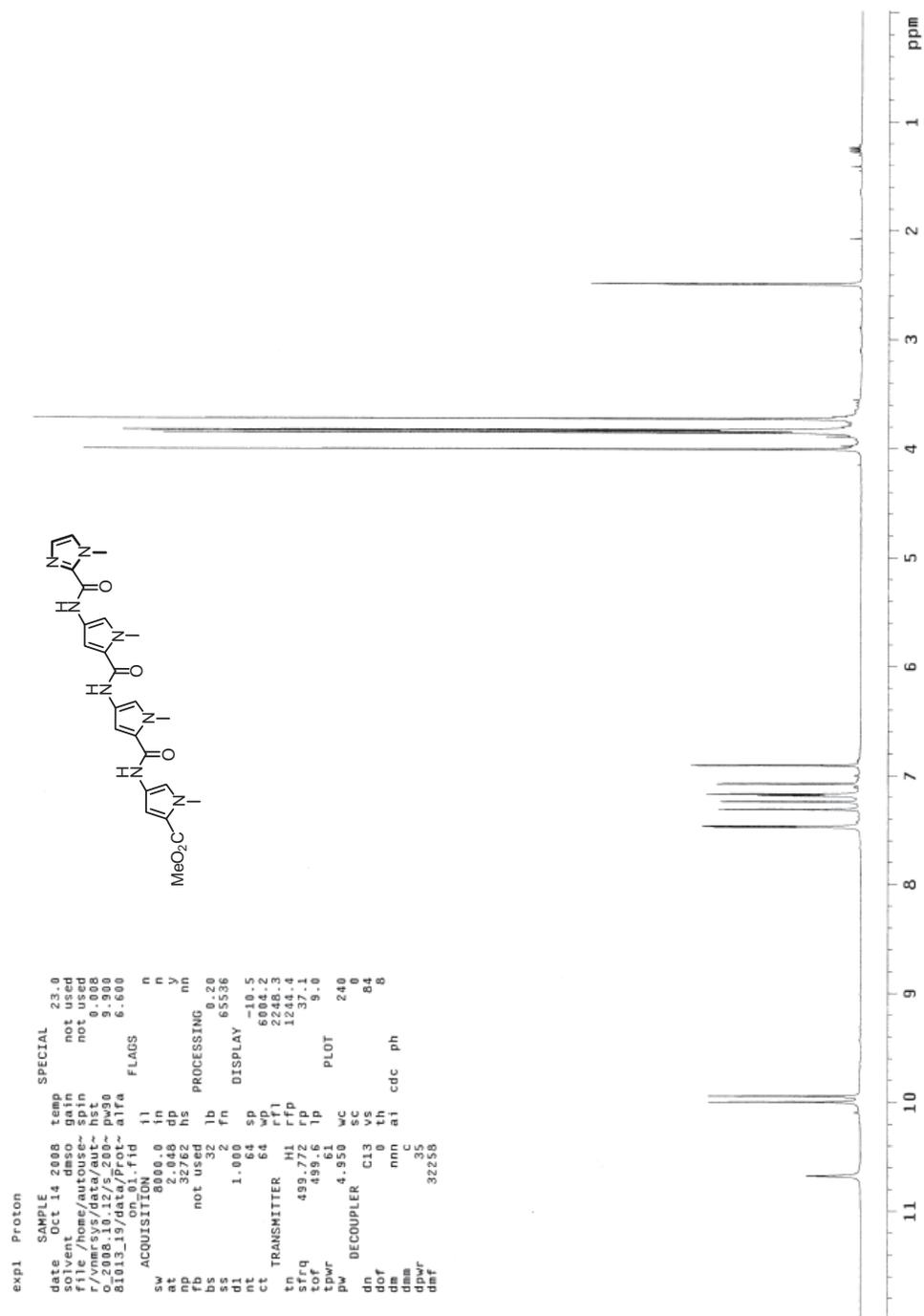


Figure 2.15 <sup>1</sup>H NMR of ImPyPyPy-CO<sub>2</sub>Me (14)

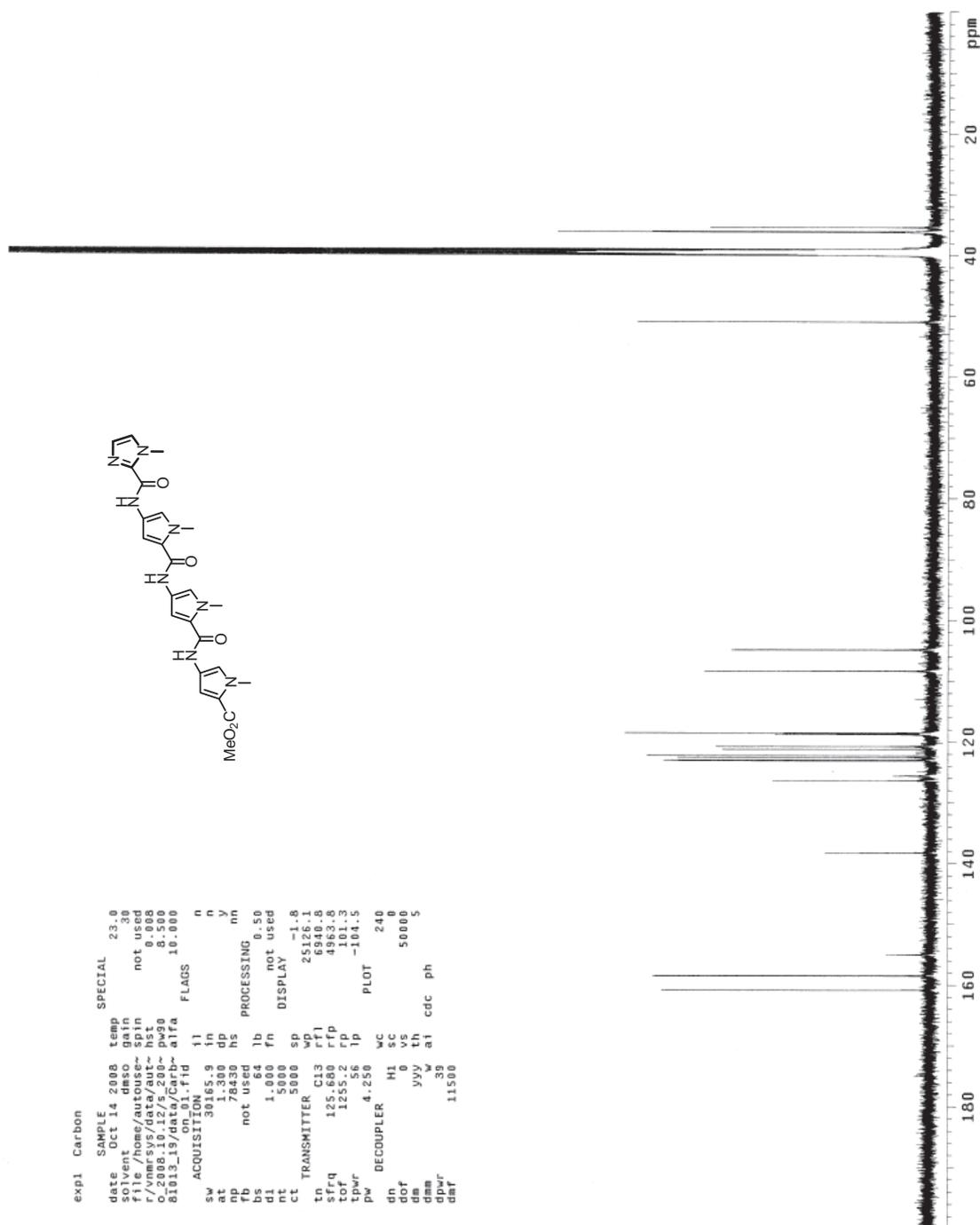


Figure 2.16 <sup>13</sup>C NMR of ImPyPyPy-CO<sub>2</sub>Me (14)

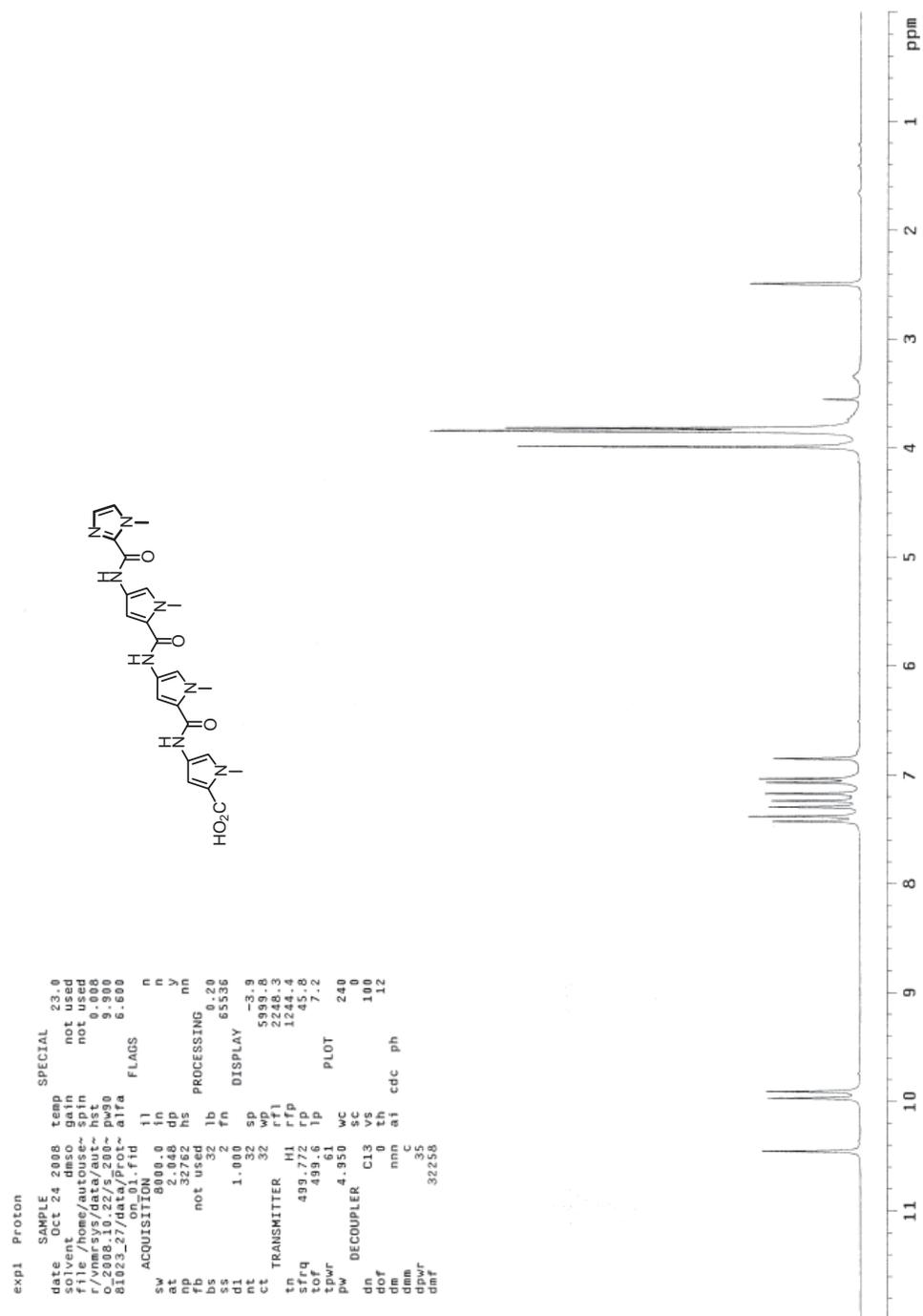
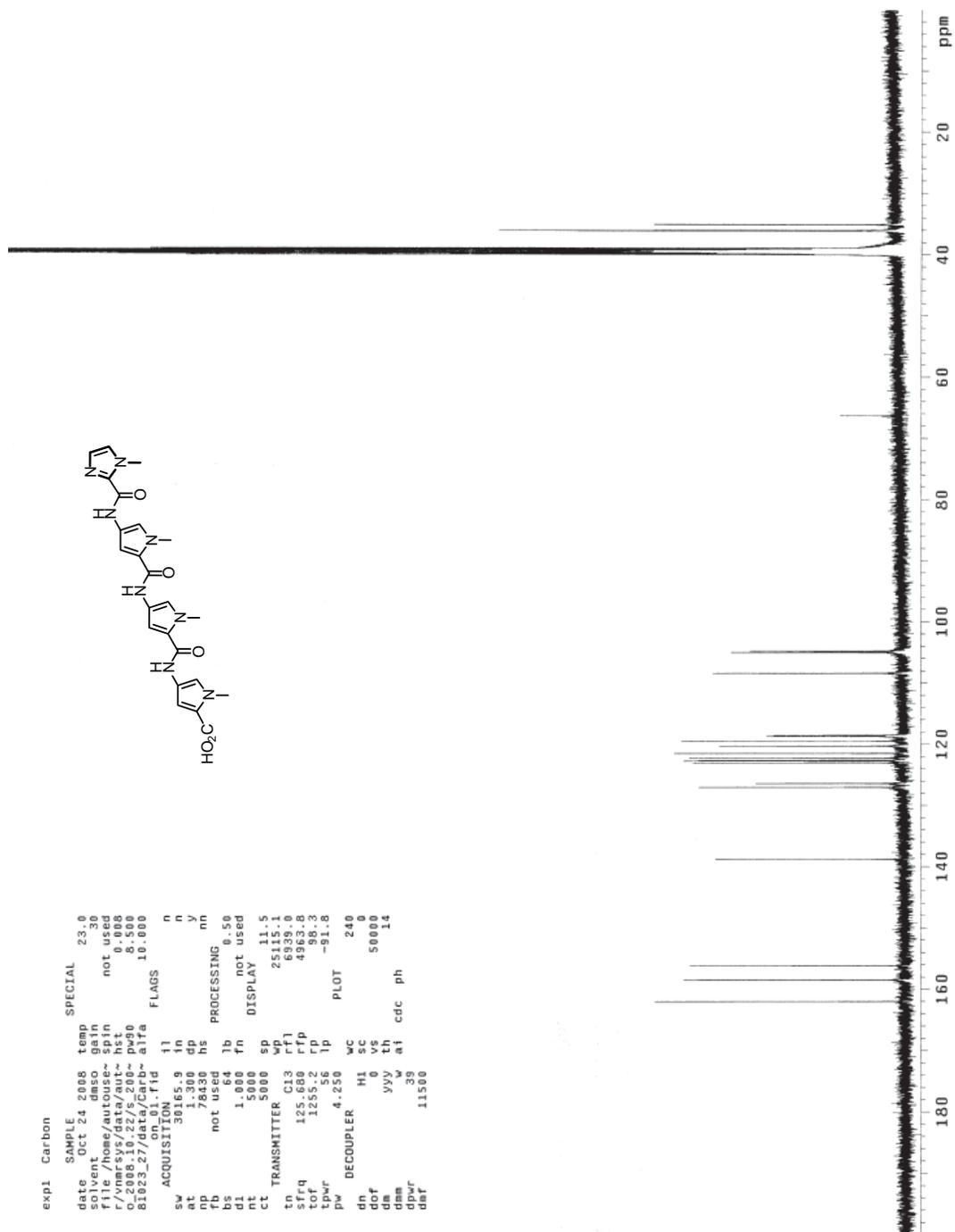


Figure 2.17  $^1\text{H}$  NMR of ImPyPyPy-CO<sub>2</sub>H (3)



**Figure 2.18** <sup>13</sup>C NMR of ImPyPyPy-CO<sub>2</sub>H (3)

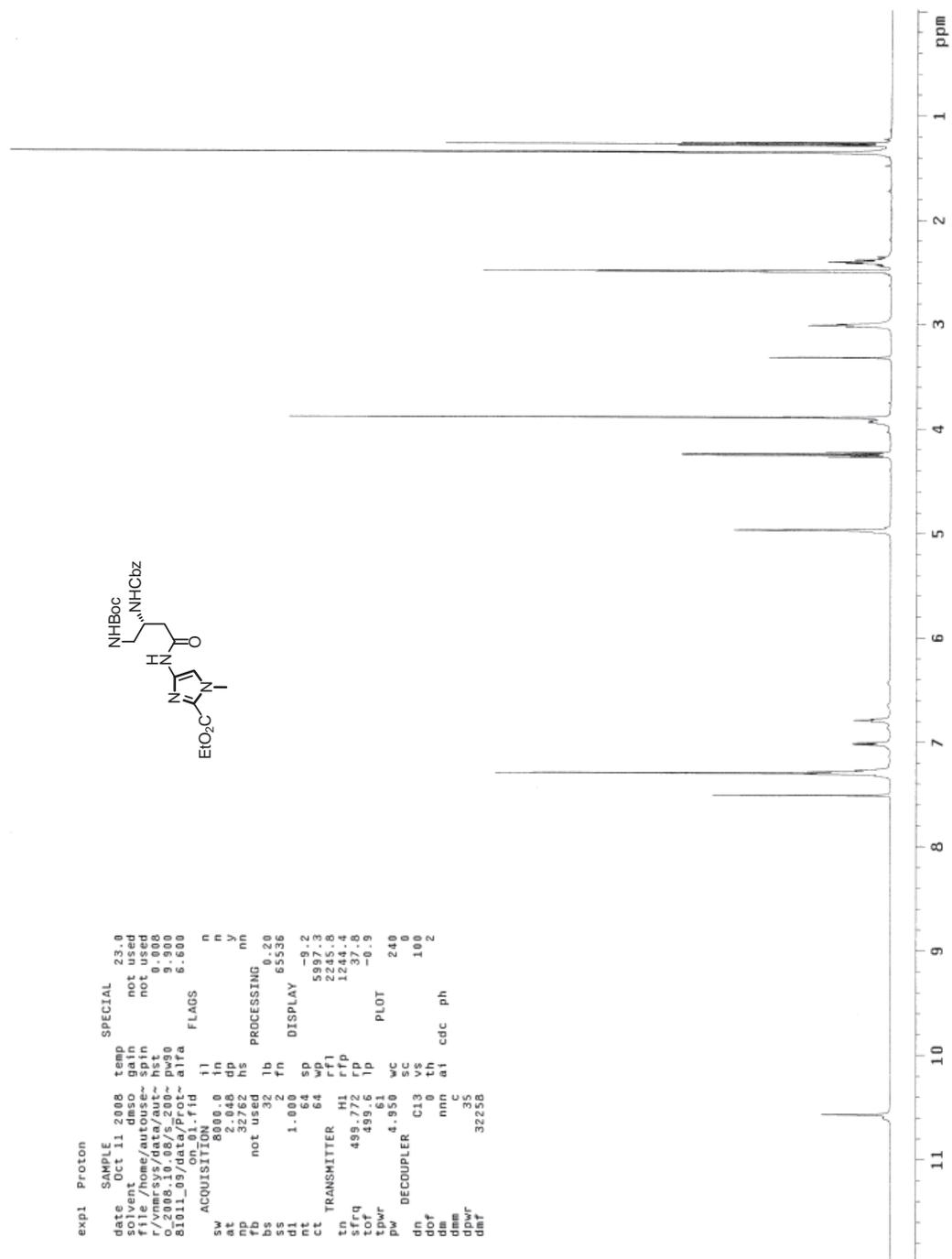
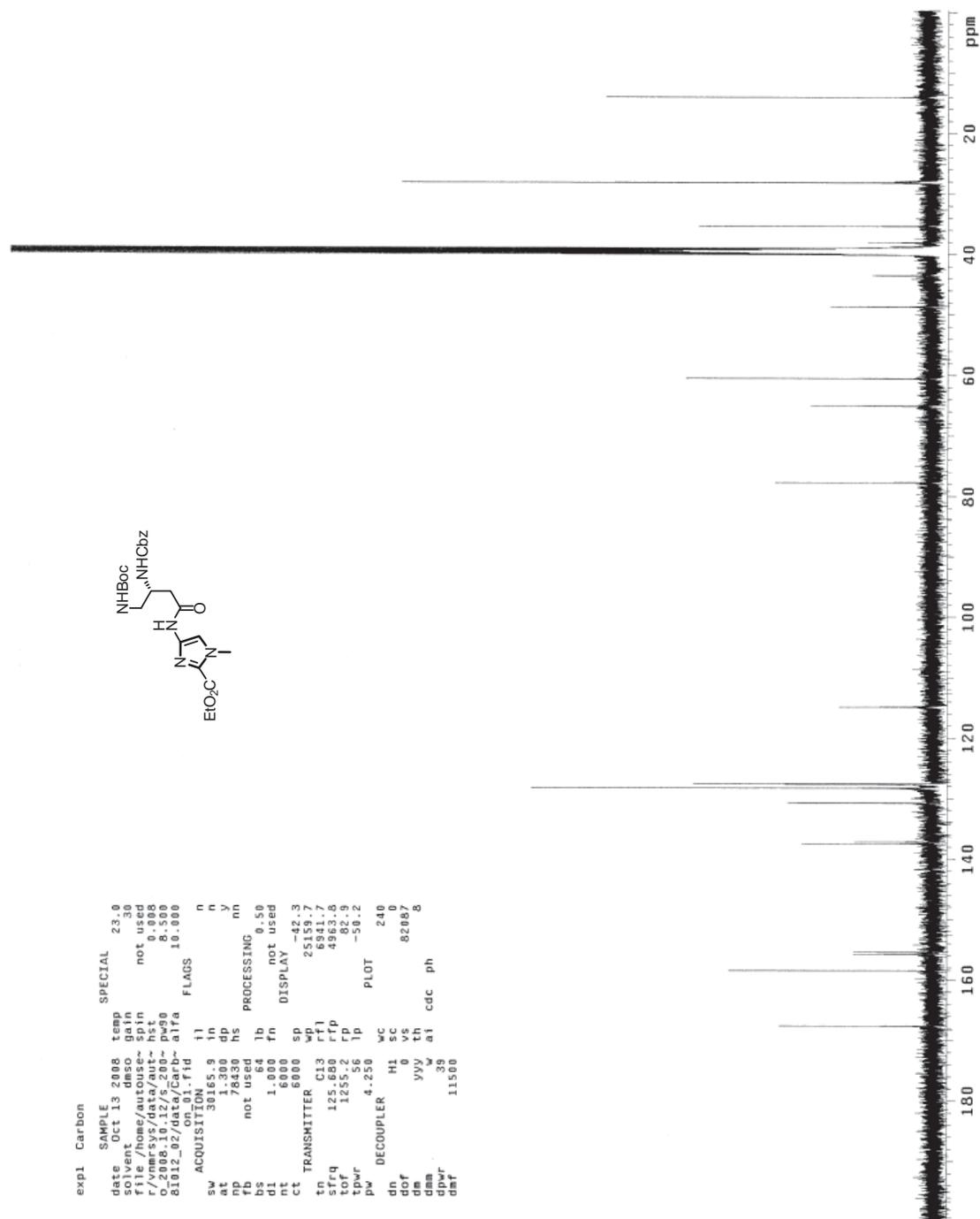
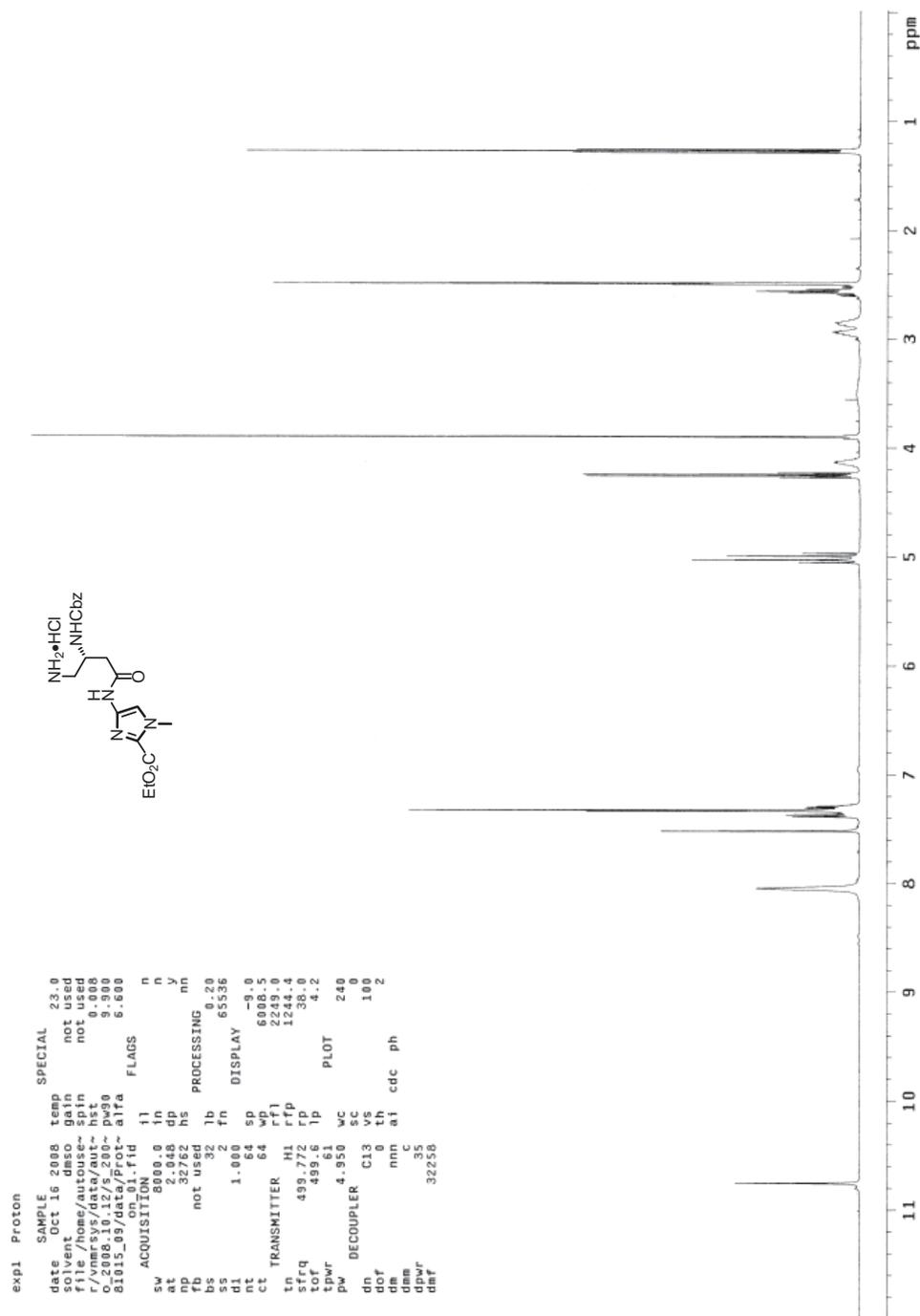


Figure 2.19  $^1\text{H}$  NMR of BocHN-(*R*) $^{\beta}$ -CbzHN- $\gamma$ -Im-CO<sub>2</sub>Et (15)



**Figure 2.20**  $^{13}\text{C}$  NMR of BocHN-(*R*) $\beta$ -CbzHN $\gamma$ -Im-CO<sub>2</sub>Et (**15**)



**Figure 2.21**  $^1\text{H}$  NMR of  $\text{HCl} \cdot \text{H}_2\text{N}-(R)^\beta\text{-CbzHN-}\gamma\text{-Im-CO}_2\text{Et}$  (**5**)

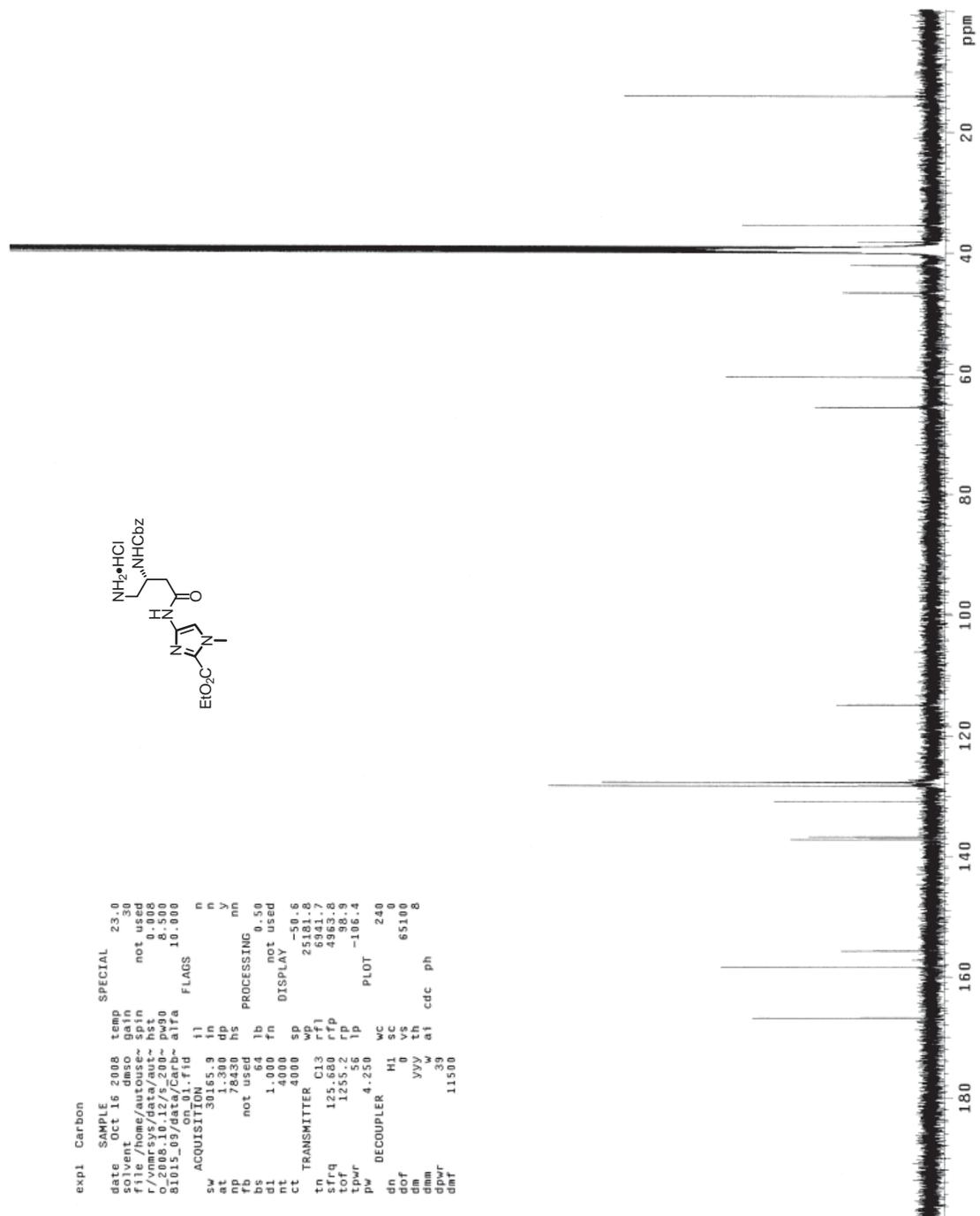
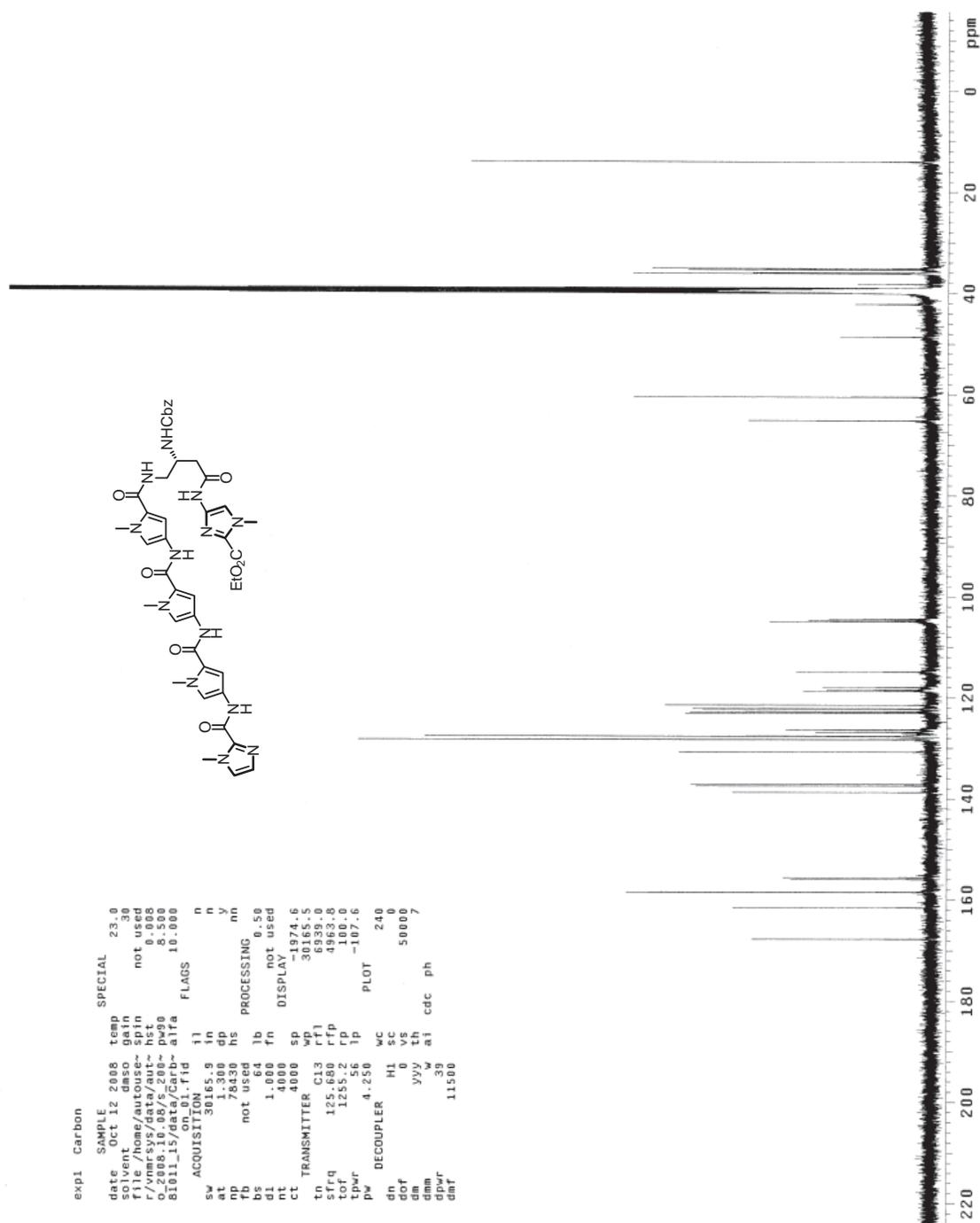


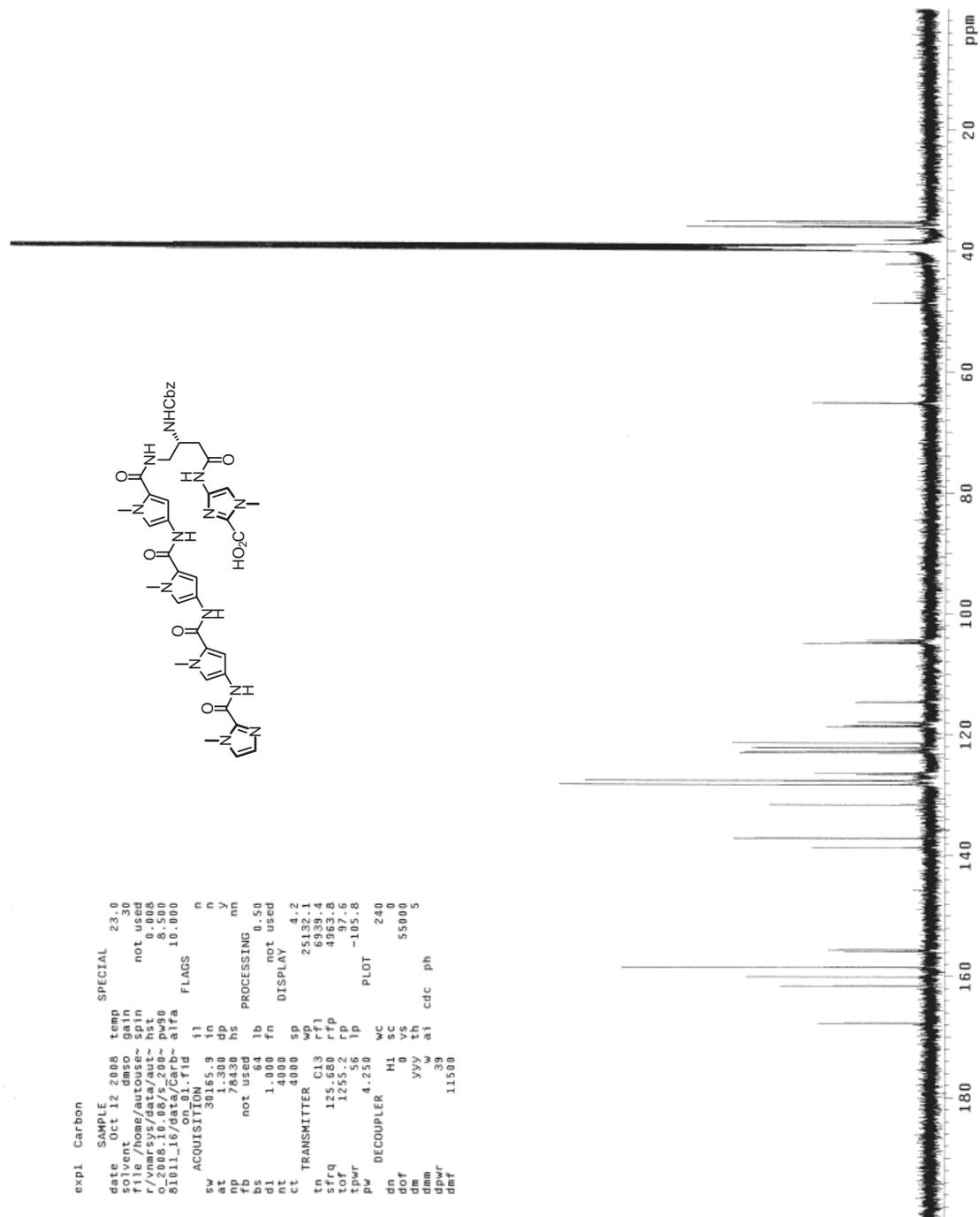
Figure 2.22  $^{13}\text{C}$  NMR of  $\text{HCl}\cdot\text{H}_2\text{N}-(R)^\beta\text{-CbzHN}\gamma\text{-Im-CO}_2\text{Et}$  (5)



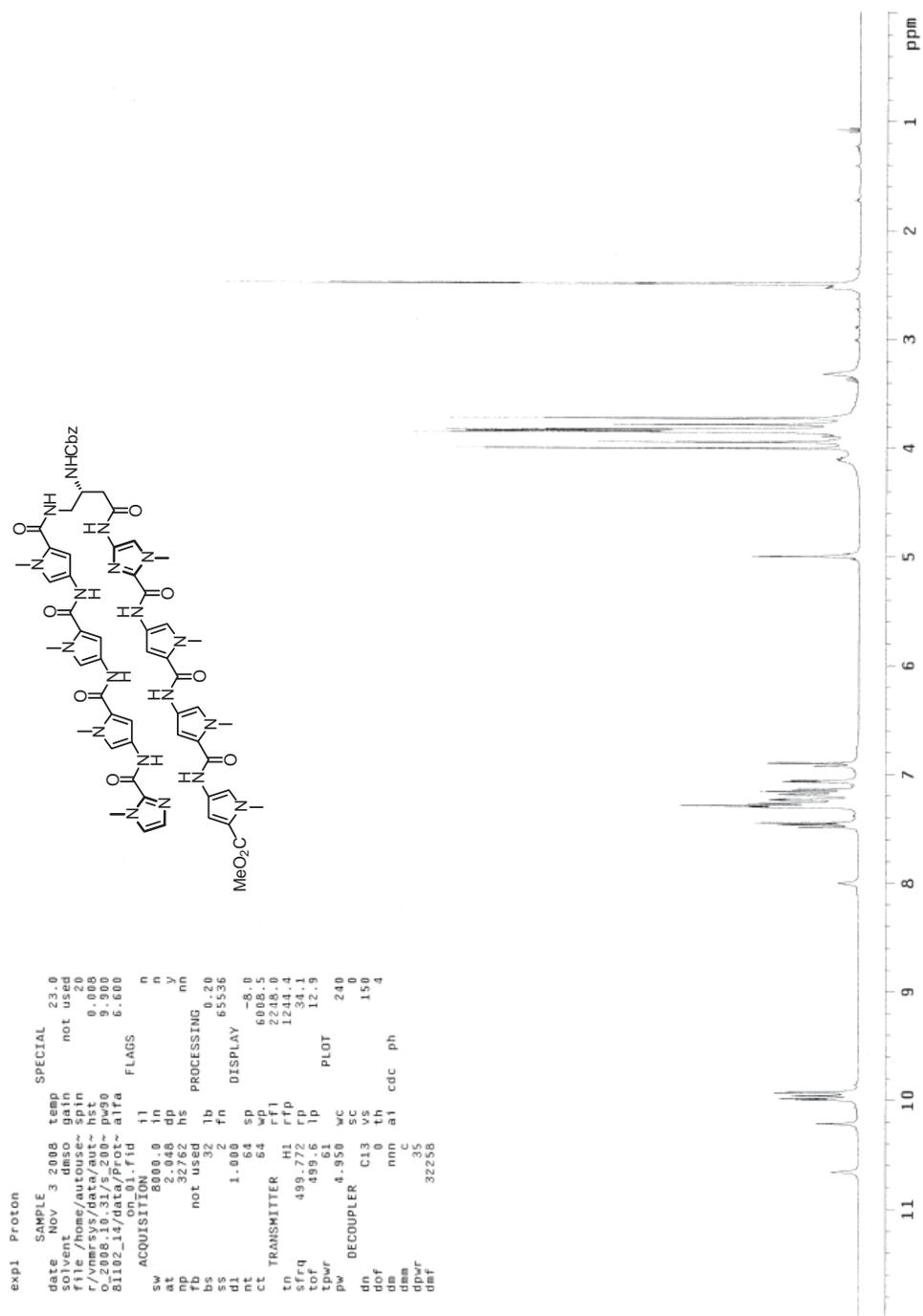


**Figure 2.24** <sup>13</sup>C NMR of ImPyPyPy-(R) $\beta$ -CbzHN $\gamma$ -Im-CO<sub>2</sub>Et (16)

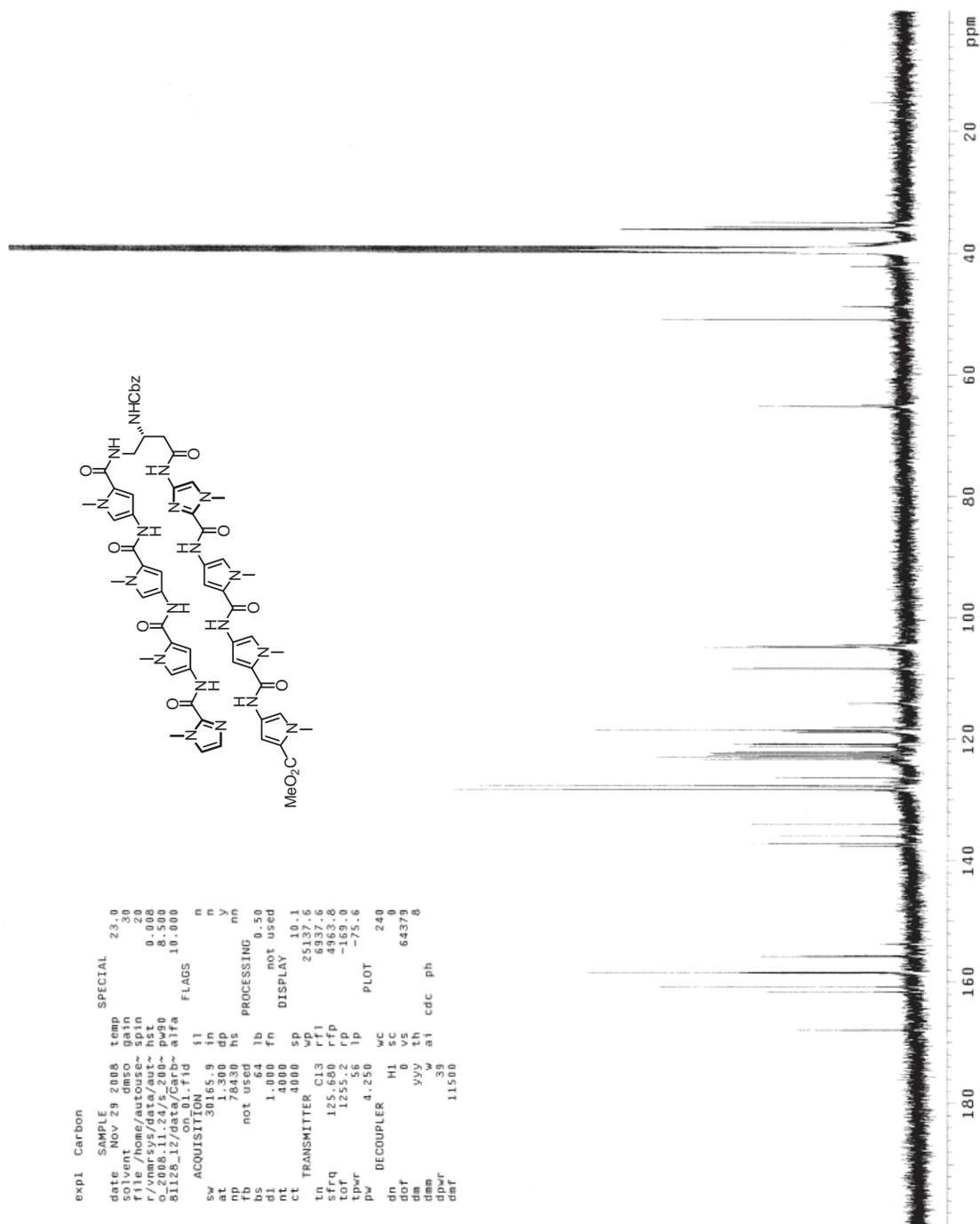




**Figure 2.26** <sup>13</sup>C NMR of ImPyPyPy-(R)- $\beta$ -CbzHN- $\gamma$ -Im-CO<sub>2</sub>H (17)



**Figure 2.27** <sup>1</sup>H NMR of ImPyPyPy-(R)<sup>β</sup>-CbzHN-γ-ImPyPyPy-CO<sub>2</sub>Me (2)



**Figure 2.28**  $^{13}\text{C}$  NMR of ImPyPyPy-(*R*) $^{\beta}$ -CbzHN $^{\gamma}$ -ImPyPyPy-CO $_2$ Me (**2**)





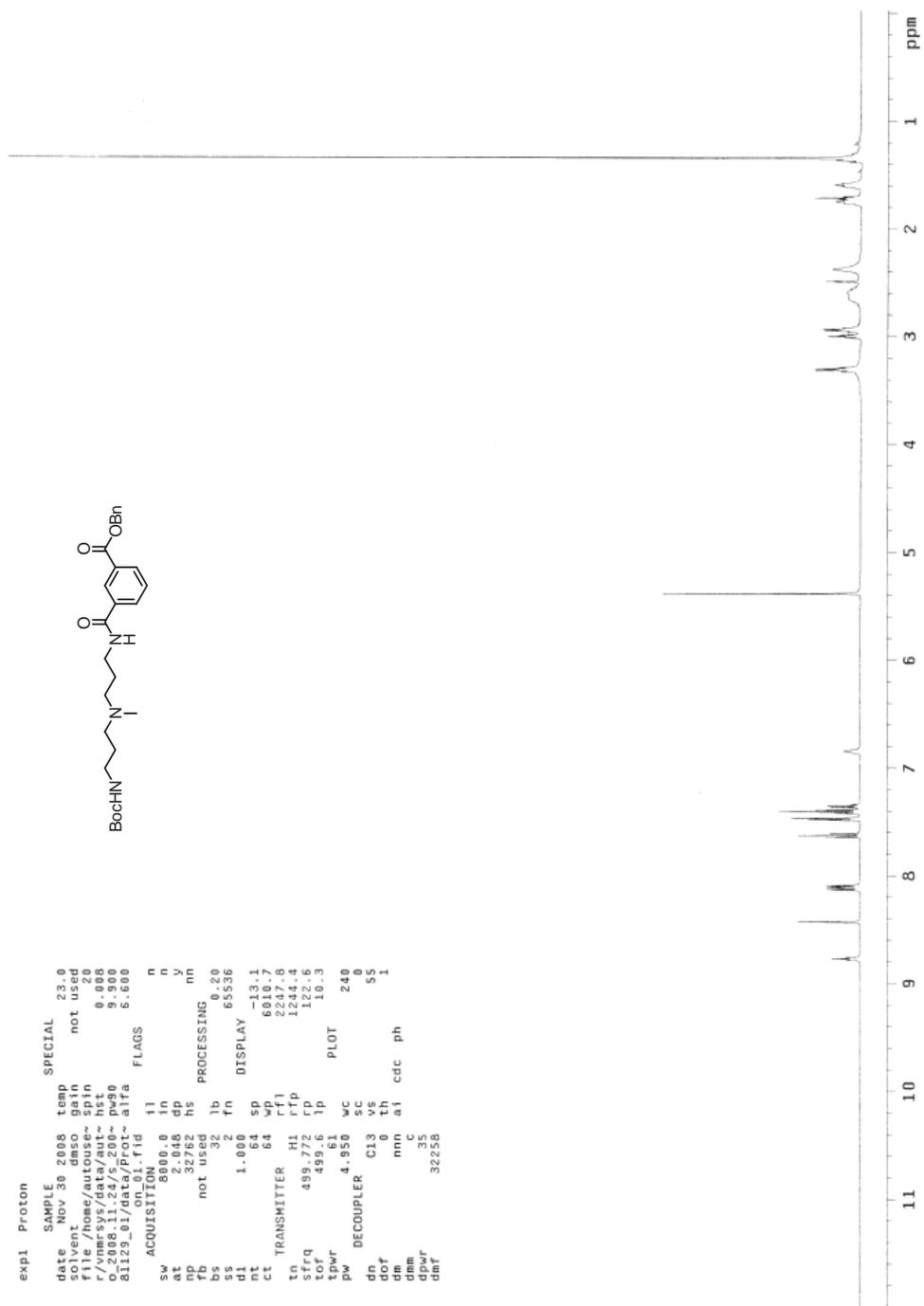
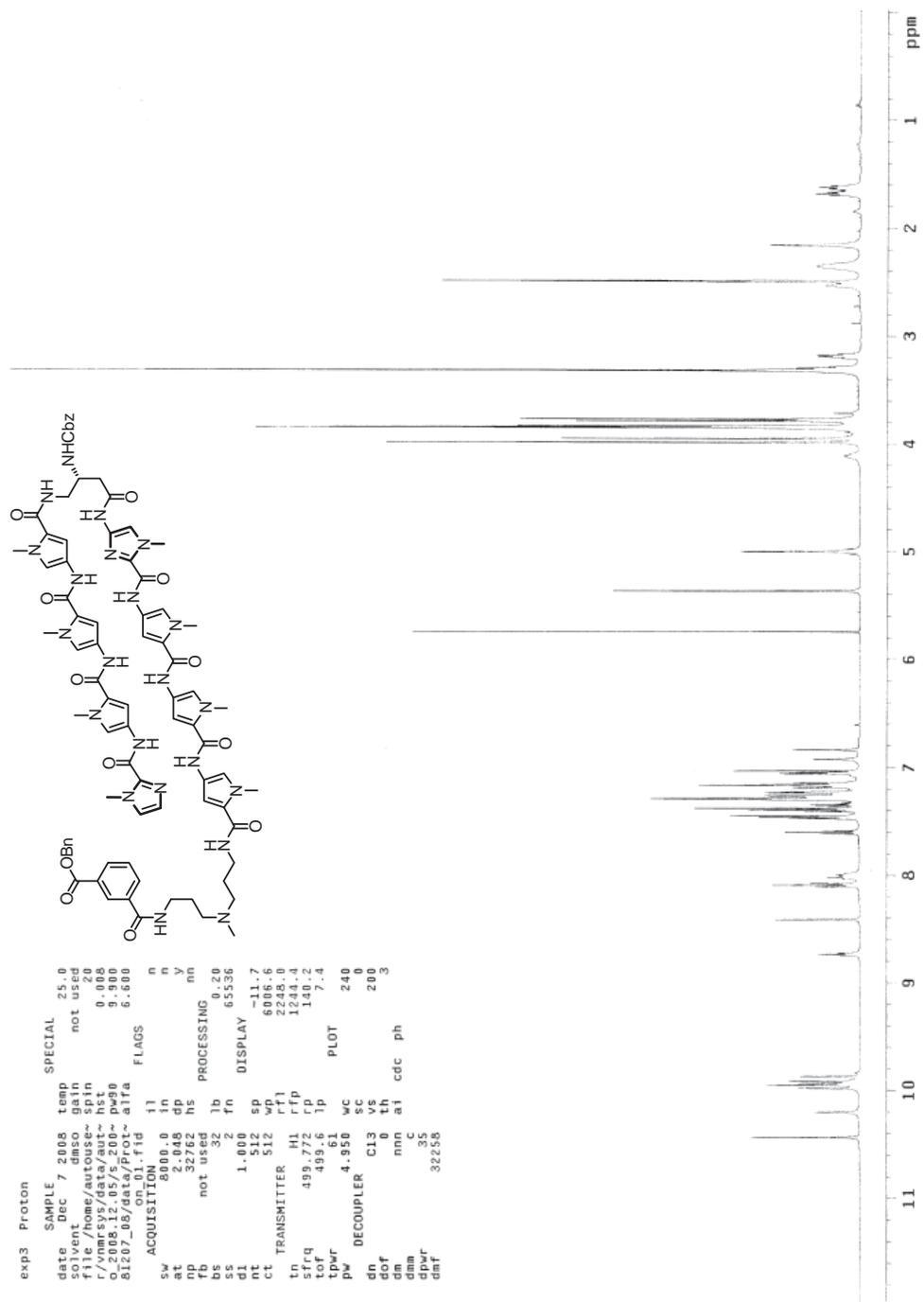


Figure 2.31 <sup>1</sup>H NMR of BoCHN-(-)-<sup>Bn</sup>OIPA (20)





**Figure 2.33**  $^1\text{H}$  NMR of ImPyPyPy-(R)<sup>β</sup>-CbzHN-γ-ImPyPyPy-(+)-BnOIPA (23)



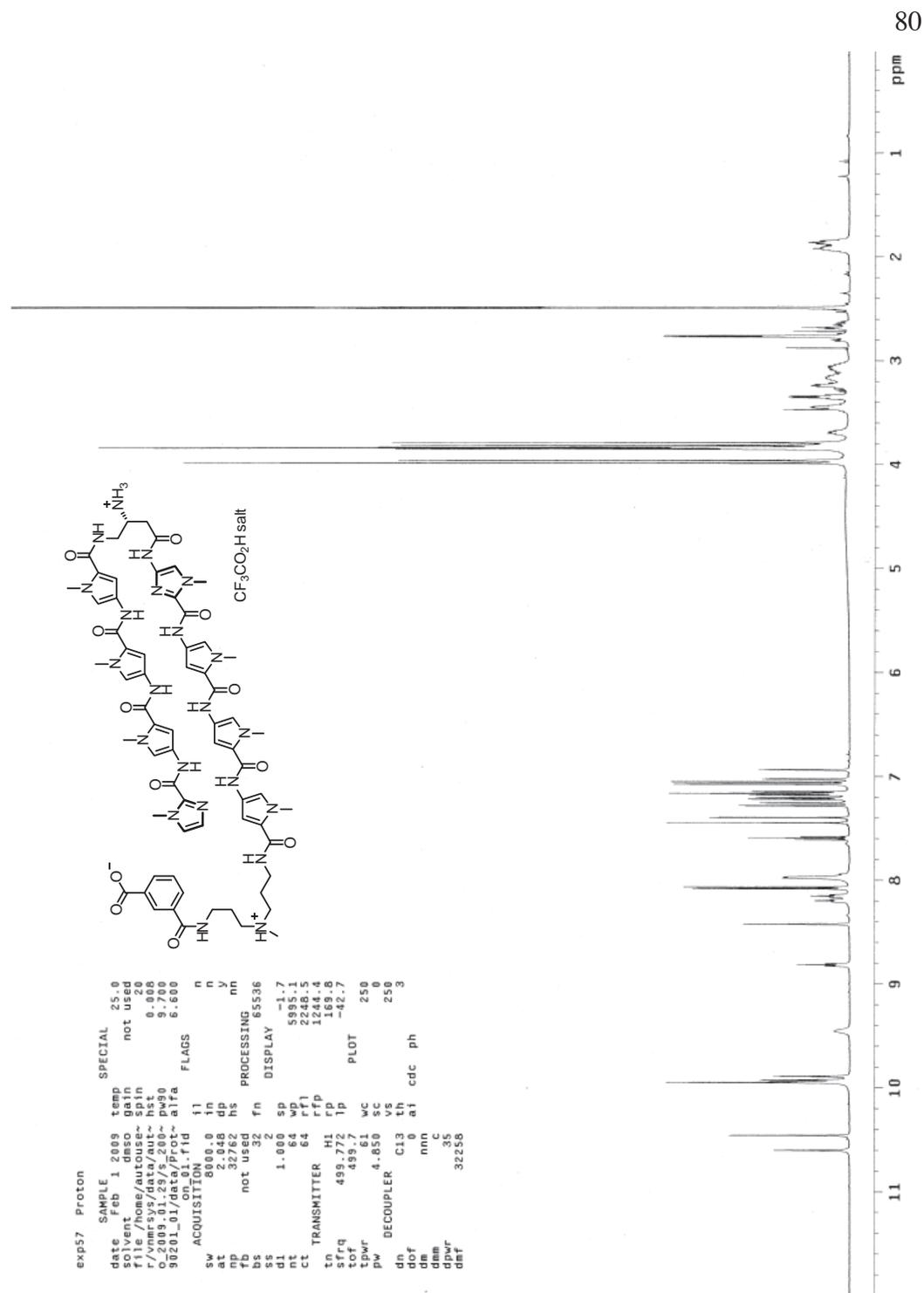
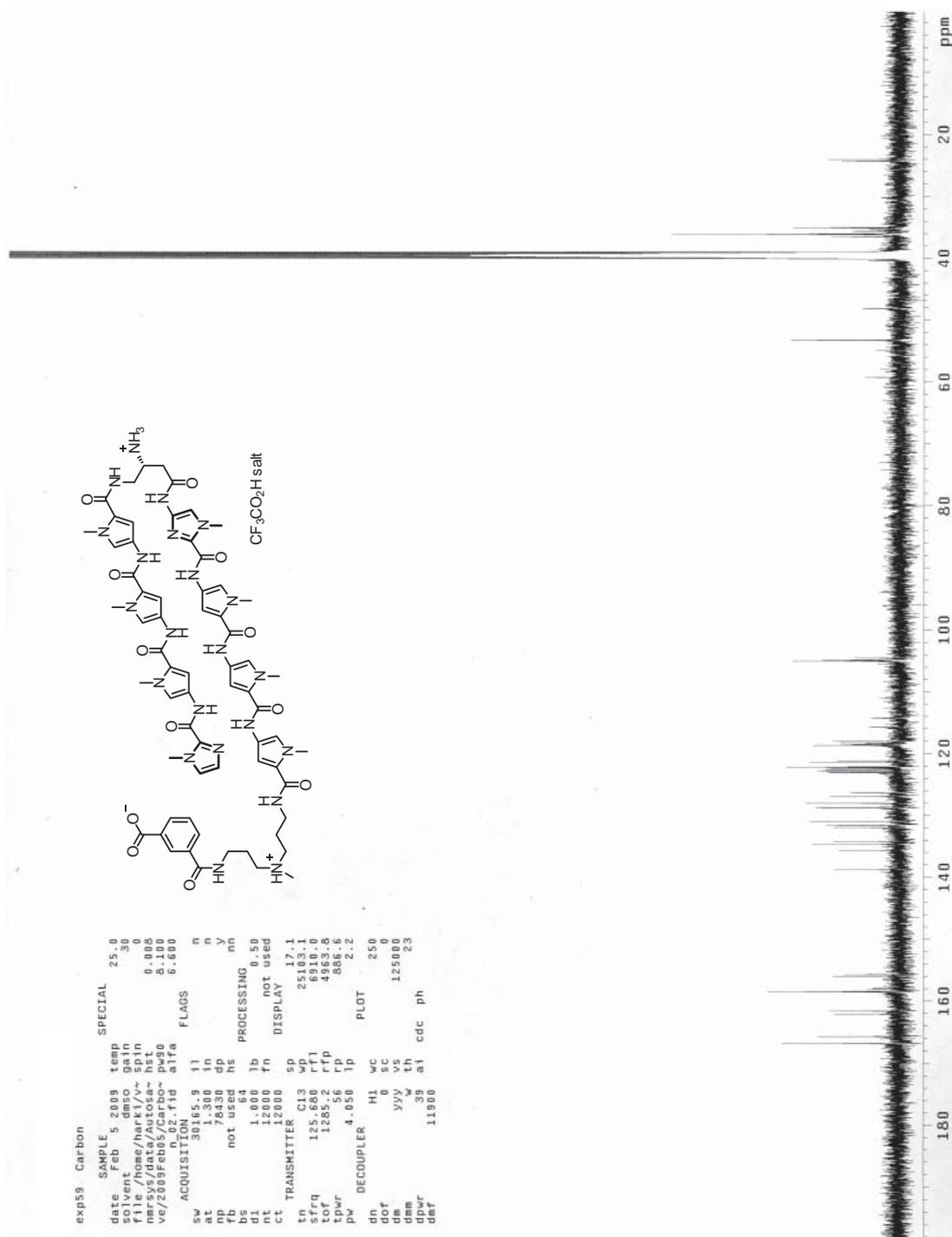


Figure 2.35 <sup>1</sup>H NMR of ImPyPyPy-(R)<sup>β</sup>-H<sub>2</sub>N<sup>γ</sup>-ImPyPyPy-(+)-IPA (1)



**Figure 2.36**  $^{13}\text{C}$  NMR of ImPyPyPy-(*R*) $^{\beta}\text{-H}_2\text{N}^{\gamma}$ -ImPyPyPy-(+)-IPA (1)

## **Chapter 3: Cyclic Pyrrole–Imidazole Polyamides Targeted to the Androgen Response Element**

*The text of this chapter was taken in part from a manuscript coauthored with Daniel A. Harki, John W. Phillips, Christian Dose, and Peter B. Dervan\* (Caltech)*

(Chenoweth, D.M., Harki, D.A., Phillips, J.W., Dose, C., Dervan, P. B. *J. Am. Chem. Soc.* **2009** In Press)

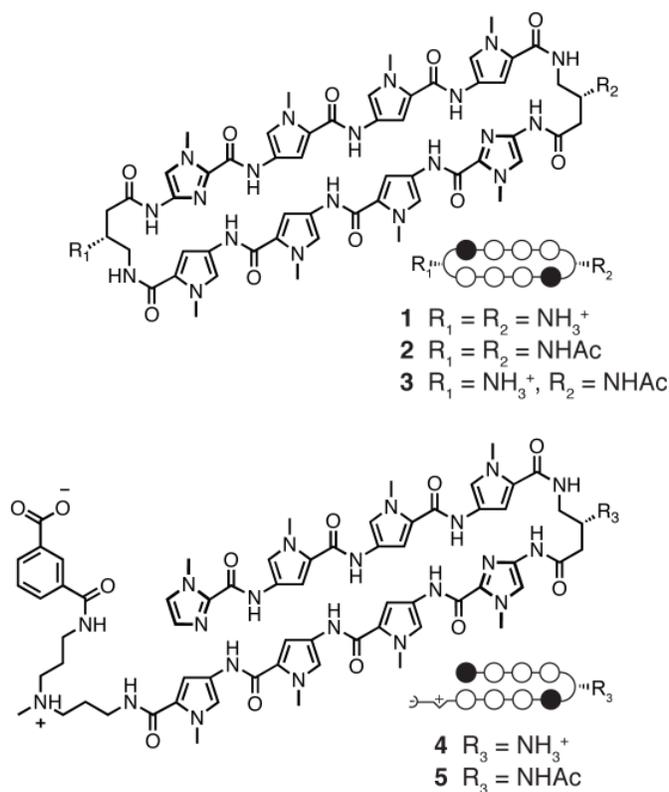
**Abstract**

Hairpin pyrrole–imidazole (Py-Im) polyamides are a class of cell-permeable DNA-binding small molecules that can disrupt transcription factor–DNA binding and regulate endogenous gene expression. The covalent linkage of antiparallel Py-Im ring pairs with an  $\gamma$ -amino acid turn unit affords the classical hairpin Py-Im polyamide structure. Closing the hairpin with a second turn unit yields a cyclic polyamide, a lesser-studied architecture mainly attributable to synthetic inaccessibility. We have applied our methodology for solution-phase polyamide synthesis to cyclic polyamides with an improved high-yield cyclization step. Cyclic 8-ring Py-Im polyamides **1–3** targets the DNA sequence 5'-WGWWCW-3' which corresponds to the androgen response element (ARE) bound by the androgen receptor transcription factor to modulate gene expression. We find that cyclic Py-Im polyamides **1–3** bind DNA with exceptionally high affinities and regulate the expression of AR target genes in cell culture studies, from which we infer that the cycle is cell permeable.

### 3.1 Introduction

Modulating the expression of eukaryotic gene networks by small molecules is a challenge at the frontier of chemical biology. Pyrrole–imidazole polyamides are a class of cell-permeable small molecules that bind to the minor groove of DNA in a sequence-specific manner.<sup>1,2</sup> Side-by-side stacked *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamides (Im/Py pairs) distinguish G•C from C•G base pairs, whereas Py/Py pairs specify for both T•A and A•T.<sup>3</sup> Py-Im hairpin polyamides have been programmed for a broad repertoire of DNA sequences with affinities similar to endogenous transcription factors.<sup>4</sup> They are cell permeable and influence gene transcription by disrupting protein–DNA interfaces.<sup>2,5,6</sup> Hairpin polyamide interference of DNA binding by transcription factors such as HIF-1 $\alpha$ ,<sup>7</sup> androgen receptor (AR),<sup>8</sup> and AP-1<sup>9</sup> has been described in recent years, yielding a new approach toward gene control by small molecules.

In parallel with our gene regulation studies, a significant effort has been devoted to maximizing the biological potency of hairpin Py-Im polyamides through structural modifications. In particular, we have recently demonstrated that hairpin polyamides bearing the (*R*)- $\beta$ -amino- $\gamma$ -turn, such as polyamide **4**, possess favorable binding affinities to DNA and are useful in gene regulation studies (Figure 3.1).<sup>5g</sup> A significant effort exists in our laboratory to regulate aberrant AR-activated gene expression in prostate cancer.<sup>8</sup> To further optimize lead oligomer **4**, it would seem reasonable that closing the hairpin with an identical linker, yielding a cyclic structure **1**, would further enhance DNA affinity (Figure 3.1). Previous syntheses of cyclic polyamides using



**Figure 3.1** Structures of cyclic and hairpin polyamides **1–5** targeted to the DNA sequence 5'-WGWWCW-3' and their ball-and-stick models. Ball-and-stick representation legend: black and white circles represent *N*-methylimidazole and *N*-methylpyrrole units, respectively, half-circle with - sign represents the terminal isophthalic acid substituent, and white half-diamond with + sign represents the triamine linker unit.

solid-phase protocols are characterized by low reaction yields due to inefficient macrocyclization.<sup>10</sup> We report here the solution-phase synthesis of cyclic polyamides **1–3** with an improved high-yield cyclization step. In addition, we examined the DNA binding properties of these compounds by thermal duplex DNA melting and performed preliminary studies of their *in vitro* ADMET properties. Cyclic Py-Im polyamides **1–3** were shown to regulate endogenous gene expression in cell culture experiments.

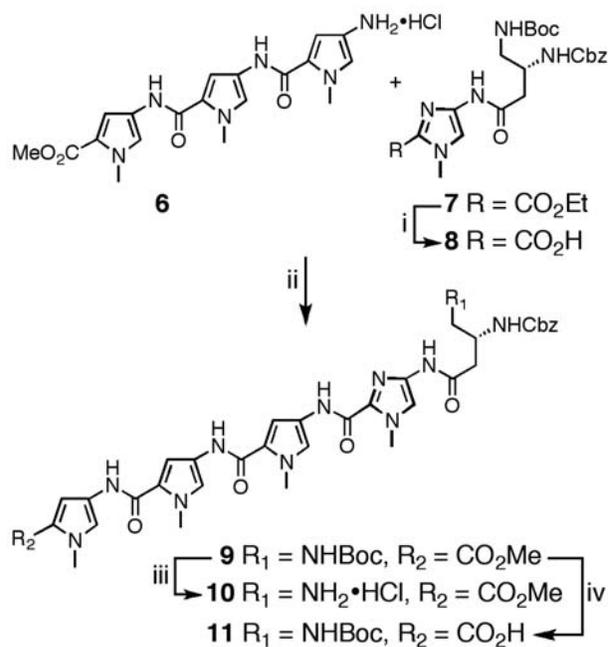
## 3.2 Results and Discussion

### 3.2.1 Solution-Phase Synthesis of Cyclic Polyamides

Due to the symmetrical nature of cyclic polyamides **1–3** and their sequence similarity to previously described hairpin polyamide **4**,<sup>11</sup> PyPyPy trimer **6** and Im-turn dimer **7** provide all the necessary atoms to synthesize **1–3**. The preparation of advanced intermediates **6** and **7** has been detailed in the Chapter 2 (this thesis)<sup>11</sup> from readily available building blocks.<sup>12</sup> The cornerstone of our synthesis strategy capitalizes on the disparate physical properties of starting materials versus products, which permit purification of most intermediates to be achieved by combinations of precipitation, trituration, and crystallization.

In addition, *in situ* deprotection of advanced pentafluorophenyl ester polyamide **14** at high dilution leads to macrocyclization in high yield, affording cyclic polyamide **15**.

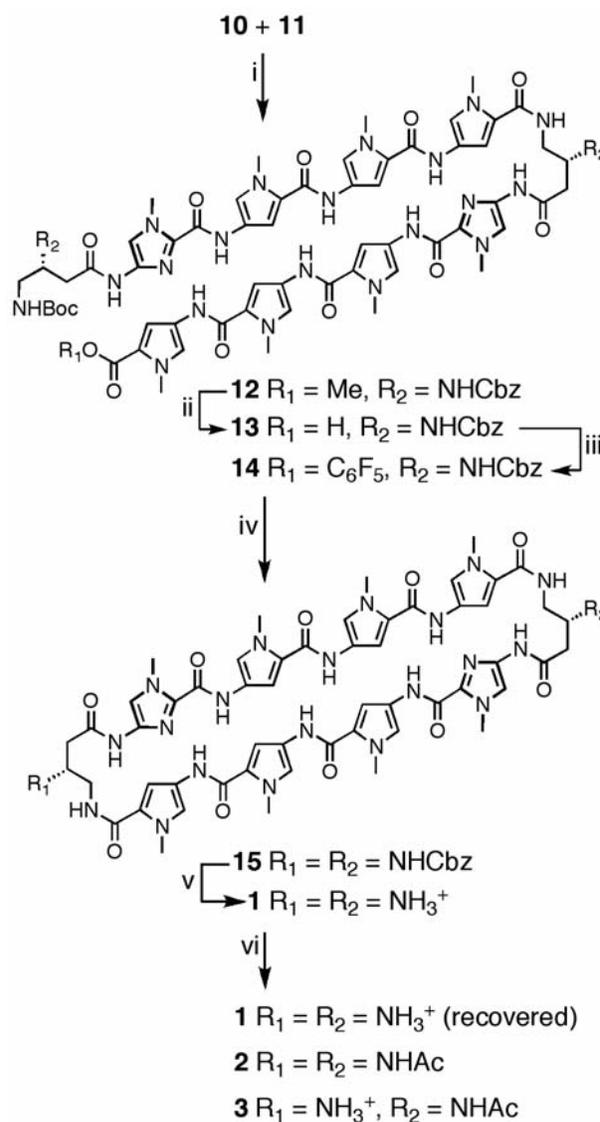
The synthesis of tetramer-turn **9** begins with Im-turn dimer **7** (Scheme 3.1). Saponification of **7** with aqueous KOH in methanol at 37 °C, followed by neutralization, precipitation, and Et<sub>2</sub>O trituration, yields Im-turn acid **8** in 95% yield. Amide coupling of **8** with pyrrole trimer **6** provides pentamer **9** in 96% yield. The utilization of a small excess of **6** relative to **8** drives the reaction to completion, and residual **6** is readily separated from **9** following precipitation in water and aqueous washing of residual solid



**Scheme 3.1** Preparation of **10** and **11**. Reagents and Conditions: (i) KOH (aq), MeOH, 37 °C, 2 h, 95%; (ii) **8**, PyBOP, DMF, DIEA, **6**, 23 °C, 4 h, 96%; (iii) HCl in 1,4-dioxane, 23 °C, 2 h, 99% (iv) NaOH (aq), 1,4-dioxane, 42 °C, 3 h, 95%.

**9.** With all atoms in place for the target cyclic polyamide **1**, compound **9** was elaborated to amine salt **10** (99% yield) by reaction with HCl in 1,4-dioxane. Carboxylic acid **11** was generated by saponification of **9** with NaOH in 1,4-dioxane in 95% yield.

Assembly of the acyclic advanced intermediate **12** was achieved by PyBOP-mediated coupling of intermediates **10** and **11** in 94% yield (Scheme 3.2). A small excess of amine salt **10** was utilized to drive the reaction to completion. Saponification of ester **12** proceeded smoothly with aqueous NaOH in 1,4-dioxane, yielding **13** in 93% yield. Activation of acid **13** as the pentafluorophenol ester **14** provided the necessary functionality to afford macrocyclization following removal of the terminal *tert*-butyl carbamate (Boc) protecting group. In our hands, we found that the pentafluorophenol ester sufficiently activated the terminal acid for amide coupling while avoiding undesired oligomerization and/or decomposition processes that are conceivable with more reactive functionalities, such as acid chlorides. Premature initiation of the macrocyclization reaction was tempered by keeping the terminal amine protonated until it was transferred into a dilute solution of acetonitrile. Addition of an amine base (DIEA) then generated the free terminal amine, which could then undergo macrocyclization in dilute solvent conditions to deliver **15**, which was directly deprotected following purification. The benzyl carbamate protecting



**Scheme 3.2** Preparation of **1**, **2**, and **3**. Reagents and Conditions: (i) PyBOP, DMF, DIEA, 23 °C, 2 h, 94%; (ii) NaOH (aq), 1,4-dioxane, 40 °C, 4 h, 93%; (iii)  $\text{CH}_2\text{Cl}_2$ , DCC, pentafluorophenol, DMAP, 23 °C, 12 h, 80%; (iv) a)  $\text{CF}_3\text{CO}_2\text{H}$ ,  $\text{CH}_2\text{Cl}_2$ , 23 °C, concentrate; b) DMF, acetonitrile, DIEA, 0–23 °C, 3 days; (v)  $\text{CF}_3\text{SO}_3\text{H}$ ,  $\text{CF}_3\text{CO}_2\text{H}$ , 23 °C, 5 min, 68% over 3 steps; (vi) NMP, DIEA,  $\text{Ac}_2\text{O}$ , 23 °C, 18% of **1** (recovered), 22% of **2**, 40% of **3**.

groups were cleaved by treatment with superacid conditions (trifluoromethylsulfonic acid–trifluoroacetic acid) to provide **1** in 68% yield over three steps. Controlled acetylation of **1** by reaction with substoichiometric quantities of Ac<sub>2</sub>O in NMP/DIEA provided a statistical population of **1** (18%), **2** (22%), and **3** (40%) that were easily separable by preparative HPLC. Acetylated hairpin **5** was prepared using excess Ac<sub>2</sub>O/pyridine in 95% yield from previously reported amine hairpin **4**.<sup>11</sup>

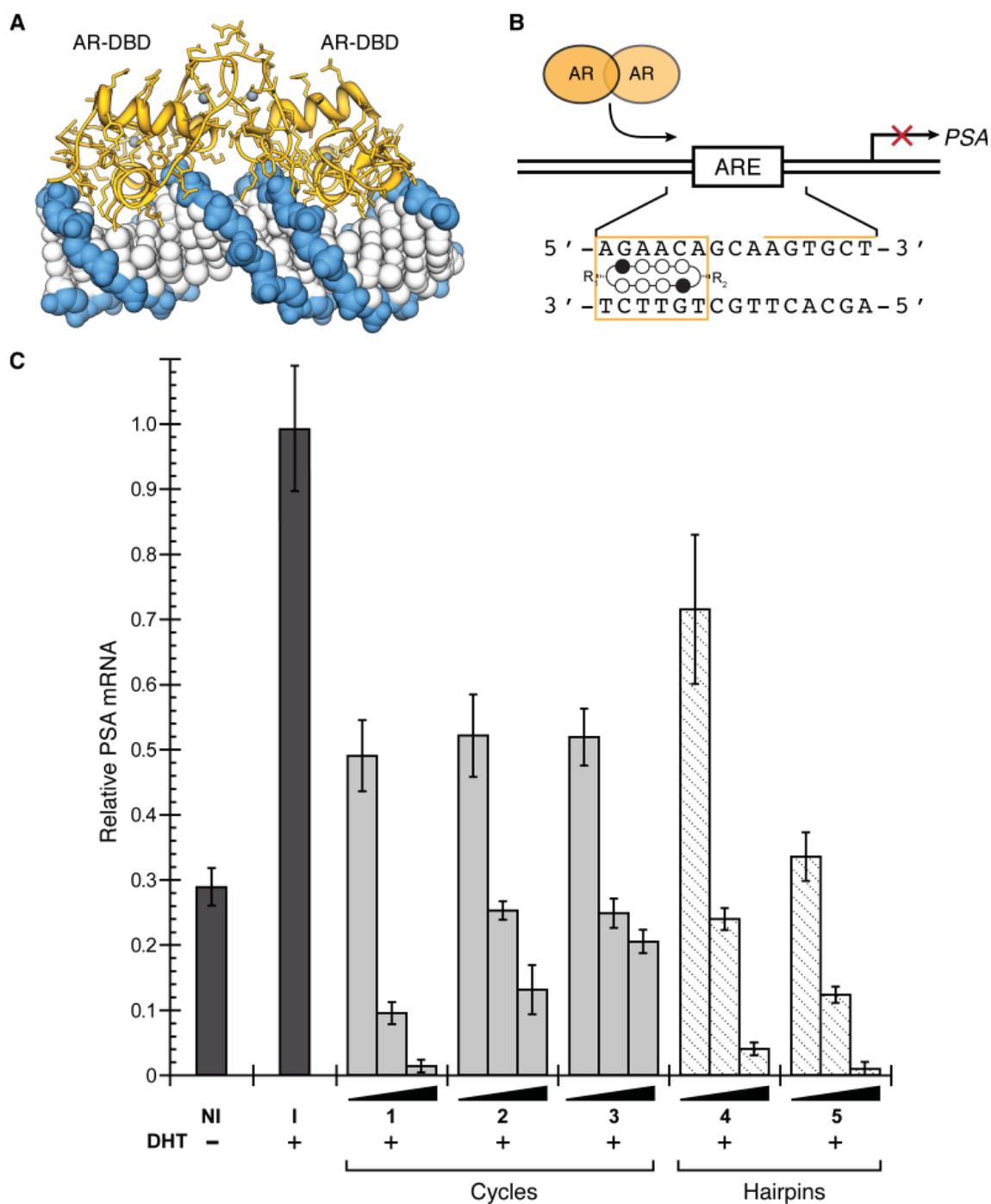
### 3.2.2 Thermal Stabilization of DNA duplexes by Polyamides

Quantitative DNase I footprint titrations have historically been utilized to measure polyamide–DNA binding affinities and specificities.<sup>13</sup> However, this method is limited to measuring  $K_a$  values  $\leq 2 \times 10^{10} \text{ M}^{-1}$ , which invalidates this technique for quantifying the exceptionally high DNA-binding affinities of cycles **1–3**.<sup>14</sup> The magnitude of DNA thermal stabilization ( $\Delta T_m$ ) of DNA–polyamide complexes has been utilized to rank order polyamides with high DNA binding affinities.<sup>5g,15</sup> Accordingly, we have employed melting temperature analysis ( $\Delta T_m$ ) for dissecting differences in DNA-binding affinities of hairpin versus cyclic polyamides. Spectroscopic analyses were performed on a 14-mer duplex DNA mimicking the androgen response element (ARE) DNA sequence, 5'-TTGCTGTTCTGCAA-3' DNA duplex, which contains one polyamide binding site. As shown in Table 3.1 polyamides **1–5** provided an increase in the duplex DNA melting temperature relative to the individual DNA duplex, thereby confirming polyamide-DNA binding. Chiral hairpin **4** led to an increased melting temperature  $\Delta T_m = 18.4 \text{ }^\circ\text{C}$  whereas cyclic polyamide **1** yielded a higher  $\Delta T_m$ -value of 23.6  $^\circ\text{C}$ . Cyclic polyamides **1–3** reveal stronger stabilizations than parent hairpin analogs **4** and **5**. Acylation of the  $\beta$ -amino turns was shown to decrease the thermal stabilization values in both hairpin and cyclic motifs, presumably due to the loss of beneficial electrostatics from the protonated-cationic amine on the turn unit.

**Table 3.1**  $T_m$  values for polyamides for **1–5**.<sup>a</sup>

ARE dsDNA sequence =		5' –TTGC TGGTCT GCAA–3'	3' –AACG ACAAGA CGTT–5'
Polyamides		$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$
–		60.0 ( $\pm 0.3$ )	–
 (1)		83.5 ( $\pm 0.5$ )	23.6 ( $\pm 0.6$ )
 (2)		81.2 ( $\pm 0.2$ )	21.3 ( $\pm 0.4$ )
 (3)		82.0 ( $\pm 0.0$ )	22.1 ( $\pm 0.3$ )
 (4)		78.4 ( $\pm 0.5$ )	18.4 ( $\pm 0.6$ )
 (5)		76.0 ( $\pm 0.5$ )	16.1 ( $\pm 0.6$ )

<sup>a</sup>All values reported are derived from at least three melting temperature experiments with standard deviations indicated in parentheses.  $\Delta T_m$  values are given as  $T_m^{\text{(DNA/polyamide)}} - T_m^{\text{(DNA)}}$ . The propagated error in  $\Delta T_m$  measurements is the square root of the sum of the square of the standard deviations for the  $T_m$  values.



**Figure 3.2** Targeting the ARE with DNA-binding polyamides. (a) X-ray crystal structure of androgen receptor homodimer DNA-binding domain bound to the sequence 5'-CTGTTCTTGATGT-TCTGG-3' (PDB 1r4i).<sup>16</sup> (b) Map of the PSA-ARE site (top) and schematic representation of a cyclic polyamide targeting the PSA-ARE site 5'-AGAACA-3'. (c) Inhibition of induced PSA mRNA expression in LNCaP cells by cyclic PI polyamides **1–3** and hairpin polyamides **4** and **5** (dosed at 0.3, 3, and 30  $\mu$ M) by real-time quantitative PCR. The results were normalized to a DHT-induced, untreated control (control=1), and the error bars represent the standard error of the mean of a single experiment performed in biological triplicate. The entire experiment was reproduced four times, with similar results. NI = noninduced, I = induced, DHT = dihydrotestosterone.

### 3.2.3 Biological Assay for Cell Permeability

Hairpin polyamides have been shown to modulate endogenous gene expression in living cells by disrupting transcription factor–DNA binding in gene promoters.<sup>2,7-9</sup> Recently, hairpin polyamide **4** was shown to inhibit androgen receptor-mediated expression of prostate-specific antigen (PSA) in LNCaP cells by targeting the DNA sequence 5′-AGAACA-3′ found in the ARE.<sup>5g</sup> We utilized this cell culture transcription assay to investigate the biological activity of cyclic polyamides **1–3** in comparison to hairpin polyamides **4** and **5**. Since small structural changes to polyamides have been shown to correlate with differences in cellular uptake properties,<sup>5</sup> it was not obvious whether cyclic polyamides **1–3** would permeate cell membranes and exhibit biological activity comparable to that of hairpin polyamides **4** and **5**. Quantitative real-time RT-PCR analysis of DHT-induced PSA expression revealed that cyclic polyamides **1–3** all decreased PSA mRNA levels in LNCaP cells, with cycle **1** exhibiting activity comparable to that of acetylated hairpin polyamide **5** (Figure 3.2). On the basis of these results, we can infer that this class of cyclic Py-Im polyamides are cell permeable and can regulate endogenous gene expression in cell culture.

### 3.2.4 ADMET Studies of Polyamides **1** and **5**

Due to the promising cell culture results obtained with cyclic polyamide **1** and hairpin polyamide **5** against PSA gene expression, we contracted preclinical *in vitro* absorption, distribution, metabolism, excretion, and toxicity (ADMET)<sup>17</sup> studies for both compounds.<sup>18</sup> Results from this study are summarized below and additional detail can be found in section 3.6 (Spectra and Supporting Information) and Appendix B of this thesis. Polyamides **1** and **5** were both found to exhibit low Caco-2 permeability, suggesting that neither compound may be orally available. Both **1** and **5** were found to be almost exclusively protein bound in plasma, with half-lives greater than 2 h. A recent positron emission tomography (PET)-based biodistribution study of a related hairpin polyamide in mice revealed high levels of liver occupancy following tail vein dosage.<sup>19</sup> On the basis of this result, we investigated the liver stability of candidate polyamides **1** and **5**. Microsomal intrinsic clearance studies found half-lives greater than 3 h for **1** and **5** in both human and rat liver microsomes, and no significant inhibition was measured against any cytochrome P450 isoform examined (Cyp1A2/CEC, Cyp2C8/DBF, Cyp2C9/DBF, Cyp2C19/DBF, Cyp2D6/AMMC, Cyp3A4/BFC, Cyp3A4/DBF). Furthermore, no obvious toxicity (IC<sub>50</sub> > 100 μM) was observed in the human hepatocellular carcinoma cell line HepG2. In addition, standard hERG FastPatch assays of cardiac toxicity found that both polyamides (**1** and **5**) were devoid of unwanted inhibition (IC<sub>50</sub>

> 100  $\mu\text{M}$ ).

### 3.3 Conclusion

We describe a solution-phase synthesis methodology for preparing cyclic Py-Im polyamides, highlighted by an efficient macrocyclization between the alkyl linker amine and a pentafluorophenol ester-activated amino acid. The three cyclic Py-Im polyamides possessed high DNA-binding affinities and were capable of accessing the nucleus in cell culture, as judged by their ability to downregulate AR-activated PSA expression in cell culture. Preclinical ADMET analysis of cyclic polyamide **1** and hairpin polyamide **5** revealed favorable drug-like properties such as high liver stability and low toxicity. Ongoing work is focused on characterizing the precise molecular interactions between cyclic polyamides and their cognate DNA sequences by high-resolution crystallographic studies.

### 3.4 Experimental Section

#### 3.4.1 General

Chemicals and solvents were purchased from Sigma-Aldrich and were used without further purification. (*R*)-3,4-Cbz-Dbu(Boc)-OH was purchased from Senn Chemicals AG (code number 44159). All DNA oligomers were purchased HPLC purified from Integrated DNA Technologies. Water (18 M $\Omega$ ) was purified using a Millipore Milli-Q purification system. Centrifugation was performed in a Beckman Coulter benchtop centrifuge (Allegra 21R) equipped with a Beckman swing-out rotor (model S4180). Analytical HPLC analysis was conducted on a Beckman Gold instrument equipped with a Phenomenex Gemini analytical column (250  $\times$  4.6 mm, 5  $\mu\text{m}$ ) and a diode array detector, and the mobile phase consisted of a gradient of acetonitrile (MeCN) in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H. Preparative HPLC was performed on an Agilent 1200 system equipped with a solvent degasser, a diode array detector, and a Phenomenex Gemini column (250  $\times$  21.2 mm, 5  $\mu\text{m}$ ). A gradient of MeCN in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H was utilized as the mobile phase. UV-Vis measurements were made on a Hewlett-Packard diode array spectrophotometer (model 8452 A), and polyamide concentrations were measured in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H using an extinction coefficient of 69200 M<sup>-1</sup>·cm<sup>-1</sup> at  $\lambda_{\text{max}}$  near 310 nm. NMR spectroscopy was performed on a Varian instrument operating at 499.8 (for <sup>1</sup>H) or 125.7 MHz (for <sup>13</sup>C) at ambient temperature. All NMR analyses were performed in DMSO-*d*<sub>6</sub>, and chemical shifts are reported in parts per million relative to the internal solvent peak referenced to 2.49 (for <sup>1</sup>H) or 39.5 (for <sup>13</sup>C). High-resolution

mass spectrometry (HRMS) was recorded in positive-ion mode by fast-atom bombardment (FAB<sup>+</sup>) on a JEOL JMS-600H instrument or by electrospray ionization (ESI<sup>+</sup>) on a Waters Acquity UPLC-LCT Premiere XE TOF-MS system.

#### 3.4.2 UV Absorption Spectrophotometry

Melting temperature analysis was performed on a Varian Cary 100 spectrophotometer equipped with a thermo-controlled cell holder possessing a cell path length of 1 cm. A degassed aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0 was used as analysis buffer. DNA duplexes and polyamides were mixed in 1:1 stoichiometry to a final concentration of 2 μM for each experiment. Prior to analysis, samples were heated to 90 °C and cooled to a starting temperature of 23 °C with a heating rate of 5 °C/min for each ramp. Denaturation profiles were recorded at λ = 260 nm from 23 to 90 °C with a heating rate of 0.5 °C/min. The reported melting temperatures were defined as the maximum of the first derivative of the denaturation profile.

#### 3.4.3 Measurement of Androgen-Induced PSA mRNA

Experiments were performed as described previously<sup>8</sup> with the following modifications: (1) all compounds and controls were prepared in neat DMSO and then diluted with media to a final concentration of 0.1% DMSO, and (2) mRNA was isolated with the RNEasy 96 kit (Qiagen, Valencia, CA).

#### 3.4.4 BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-Im-CO<sub>2</sub>H (**8**)

A solution of BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-Im-CO<sub>2</sub>Et **7** (450 mg, 0.894 mmol) dissolved in MeOH (1.0 mL) and aqueous KOH (1 N, 2.0 mL, 2.0 mmol) was stirred at 37 °C for 2 h. The reaction mixture was added to a cooled (ice bath) solution of distilled H<sub>2</sub>O (10 mL) preacidified with aqueous HCl (1 N, 2.0 mL, 2.0 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (10 mL) and collected by centrifugation. The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness and then suspended in excess anhydrous Et<sub>2</sub>O and filtered, and the filter cake washed with copious amounts of anhydrous Et<sub>2</sub>O. Drying of the brown solid *in vacuo* yielded saponified dimer **8** (404 mg, 95%). <sup>1</sup>H NMR: δ 10.46 (s, 1H), 7.47 (s, 1H), 7.31-7.26 (m, 5H), 7.02 (d, *J* = 8.3 Hz, 1H), 6.79 (t, *J* = 5.4 Hz, 1H), 4.97 (s, 2H), 3.92 (m, 1H), 3.88 (s, 3H), 3.60 (s, 1H), 3.01 (m, 2H),

2.41 (m, 2H), 1.35 (s, 9H);  $^{13}\text{C}$  NMR:  $\delta$  167.6, 160.0, 155.8, 155.4, 137.11, 137.09, 131.6, 128.2, 127.66, 127.55, 114.6, 77.7, 65.1, 48.7, 43.5, 38.0, 35.4, 28.2; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{22}\text{H}_{30}\text{N}_5\text{O}_7$  [M+H]<sup>+</sup> 476.2145, found 476.2130.

### 3.4.5 BocHN-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me (9)

A solution of BocHN-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -Im-CO<sub>2</sub>H **8** (300 mg, 0.631 mmol) and PyBOP (345 mg, 0.663 mmol) in DMF (3.2 mL) and DIEA (330  $\mu\text{L}$ , 1.9 mmol) was stirred at 23 °C for 10 min. The solution was then treated with solid (powdered) HCl•H<sub>2</sub>N-PyPyPy-CO<sub>2</sub>Me **6** (288 mg, 0.663 mmol) and stirred at 23 °C for 4 h. The solution was then added to distilled H<sub>2</sub>O (10 mL) preacidified with aqueous HCl (1 N, 2 mL, 2 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (10 mL) and collected by centrifugation (repeated 3x). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield BocHN-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me **9** as a tan solid (518 mg, 96%).  $^1\text{H}$  NMR:  $\delta$  10.17 (s, 1H), 10.00 (s, 1H), 9.95 (s, 1H), 9.93 (s, 1H), 7.46 (d,  $J$  = 1.7 Hz, 1H), 7.44 (s, 1H), 7.31-7.29 (m, 5H), 7.27 (d,  $J$  = 1.7 Hz, 1H), 7.23 (d,  $J$  = 1.7 Hz, 1H), 7.14 (d,  $J$  = 1.7 Hz, 1H), 7.07 (d,  $J$  = 1.7 Hz, 1H), 7.04 (d,  $J$  = 8.3 Hz, 1H), 6.90 (d,  $J$  = 2.0 Hz, 1H), 6.81 (t,  $J$  = 5.9 Hz, 1H), 4.98 (s, 2H), 3.96 (m, 1H), 3.95 (s, 3H), 3.85 (s, 3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.73 (s, 3H), 3.03 (m, 2H), 2.46 (m, 2H), 1.36 (s, 9H);  $^{13}\text{C}$  NMR:  $\delta$  167.8, 160.8, 158.5, 158.4, 155.84, 155.81, 155.5, 137.1, 136.0, 134.0, 128.3, 127.7, 127.6, 123.06, 123.00, 122.5, 122.2, 121.2, 120.7, 118.7, 118.6, 118.5, 114.0, 108.4, 104.9, 77.8, 65.2, 50.9, 48.8, 43.6, 38.2, 36.20, 36.18, 36.09, 34.9, 28.2; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{41}\text{H}_{49}\text{N}_{11}\text{O}_{10}$  [M]<sup>+</sup> 855.3663, found 855.3688.

### 3.4.6 HCl•H<sub>2</sub>N-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me (10)

A solution of BocHN-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me **9** (125 g, 0.146 mmol) in anhydrous HCl in 1,4-dioxane (4.0 M, 10 mL) was stirred at 23 °C for 2 h. The mixture was then diluted with 100 mL of anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield HCl•H<sub>2</sub>N-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me **10** as a brown solid (114 mg, 99%).  $^1\text{H}$  NMR:  $\delta$  10.38 (s, 1H), 9.98 (s, 1H), 9.96 (s, 1H), 9.94 (s, 1H), 8.10 (m, 3H), 7.46 (d,  $J$  = 1.7 Hz, 1H), 7.45 (s, 1H), 7.42 (d,  $J$  = 8.3 Hz, 1H), 7.34-7.28 (m, 5H), 7.28 (d,  $J$  = 1.7 Hz, 1H), 7.24 (d,  $J$  = 1.7 Hz, 1H), 7.16 (d,  $J$  = 1.7 Hz, 1H), 7.08 (d,  $J$  = 1.7

Hz, 1H), 6.90 (d,  $J = 1.7$  Hz, 1H), 5.02 (m, 2H), 4.14 (m, 1H), 3.95 (s, 3H), 3.85 (s, 3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.73 (s, 3H), 3.02 (m, 2H), 2.63 (m, 2H);  $^{13}\text{C}$  NMR:  $\delta$  167.0, 160.8, 158.5, 158.4, 155.7, 136.8, 135.7, 134.0, 128.3, 127.8, 127.7, 123.05, 122.97, 122.5, 122.2, 121.1, 120.7, 118.64, 118.60, 118.5, 108.4, 104.9, 65.6, 50.9, 46.7, 42.0, 38.2, 36.2, 36.1, 36.0, 34.9; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{36}\text{H}_{42}\text{N}_{11}\text{O}_8$  [M+H]<sup>+</sup> 756.3218, found 756.3211.

### 3.4.7 BocHN-(*R*) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>H (**11**)

A solution of BocHN-(*R*) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me **9** (200 mg, 0.234 mmol) dissolved in 1,4-dioxane (2.3 mL) and aqueous NaOH (1 N, 2.3 mL, 2.3 mmol) was stirred at 42 °C for 3 h. The solution was then added to distilled H<sub>2</sub>O (5 mL) preacidified with aqueous HCl (1 N, 2.3 mL, 2.3 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (10 mL) and collected by centrifugation (repeated 2x). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness and then suspended in excess anhydrous Et<sub>2</sub>O and filtered, and the filter cake washed with copious amounts of anhydrous Et<sub>2</sub>O. Drying of the tan solid *in vacuo* yielded BocHN-(*R*) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>H **11** (187 mg, 95%).  $^1\text{H}$  NMR:  $\delta$  12.15 (s, 1H), 10.21 (s, 1H), 10.00 (s, 1H), 9.96 (s, 1H), 9.92 (s, 1H), 7.44 (s, 1H), 7.42 (d,  $J = 1.7$  Hz, 1H), 7.31-7.29 (m, 5H), 7.28 (d,  $J = 1.5$  Hz, 1H), 7.24 (d,  $J = 1.5$  Hz, 1H), 7.16 (d,  $J = 1.5$  Hz, 1H), 7.08 (m, 2H), 6.85 (d,  $J = 1.7$  Hz, 1H), 6.82 (t,  $J = 5.7$  Hz, 1H), 4.98 (s, 2H), 3.95 (m, 4H), 3.85 (s, 3H), 3.84 (s, 3H), 3.81 (s, 3H), 3.03 (m, 2H), 2.46 (m, 2H), 1.36 (s, 9H);  $^{13}\text{C}$  NMR:  $\delta$  167.8, 162.0, 158.44, 158.38, 155.80, 155.77, 155.4, 137.1, 136.0, 133.9, 128.3, 127.7, 127.6, 123.0, 122.7, 122.6, 122.2, 121.2, 120.2, 119.5, 118.6, 118.5, 114.0, 108.4, 104.87, 104.83, 77.7, 65.1, 48.8, 43.5, 38.2, 36.2, 36.13, 36.06, 34.9, 28.2; HRMS (FAB<sup>+</sup>) calc'd for  $\text{C}_{40}\text{H}_{47}\text{N}_{11}\text{O}_{10}$  [M]<sup>+</sup> 841.3507, found 841.3498.

### 3.4.8 BocHN-(*R*) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-(*R*) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me (**12**)

A solution of BocHN-(*R*) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>H **11** (25 mg, 0.029 mmol) and PyBOP (17 mg, 0.031 mmol) in DMF (150  $\mu\text{L}$ ) and DIEA (16  $\mu\text{L}$ , 0.089 mmol) was stirred at 23 °C for 20 min. The solution was then treated with solid (powdered) HCl•H<sub>2</sub>N-(*R*) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me **10** (25 mg, 0.031 mmol) and stirred at 23 °C for 2 h. The solution was then added to distilled H<sub>2</sub>O (10 mL) preacidified with aqueous HCl (1 N, 1 mL, 1 mmol), yielding a precipitate that was isolated by centrifugation (~4500 rpm). The residual solid was again suspended in distilled H<sub>2</sub>O (10 mL) and collected by centrifugation (repeated 3x). The resultant solid, which contained a small amount of

residual H<sub>2</sub>O, was frozen and lyophilized to dryness. The solid was triturated with anhydrous Et<sub>2</sub>O and filtered over a sintered glass funnel. The resultant solid was washed with copious amounts of anhydrous Et<sub>2</sub>O and dried *in vacuo* to yield BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-CO<sub>2</sub>Me **12** as a tan solid (44 mg, 94%). <sup>1</sup>H NMR: δ 10.20 (s, 1H), 10.16 (s, 1H), 9.98 (s, 2H), 9.94-9.91 (m, 4H), 7.99 (m, 1H), 7.46 (d, *J* = 1.7 Hz, 1H), 7.45 (s, 1H), 7.44 (s, 1H), 7.32-7.14 (m, 18H), 7.07 (m, 2H), 7.03 (d, *J* = 8.3 Hz, 1H), 6.92 (s, 1H), 6.90 (d, *J* = 2.0 Hz, 1H), 6.80 (t, *J* = 5.6 Hz, 1H), 4.99 (m, 4H), 4.10 (m, 1H), 3.95 (m, 7H), 3.85-3.83 (m, 15H), 3.79 (s, 3H), 3.73 (s, 3H), ~3.30 (m, 2H, obstructed by H<sub>2</sub>O peak), 3.04 (m, 2H), 2.53-2.44 (m, 4H, partially obstructed by NMR solvent), 1.36 (s, 9H); <sup>13</sup>C NMR: δ 167.9, 167.8, 161.6, 160.8, 158.5, 158.44, 158.42, 155.8, 155.6, 155.5, 137.1, 136.0, 134.00, 133.98, 128.3, 127.7, 127.63, 127.60, 123.11, 123.07, 123.00, 122.80, 122.77, 122.5, 122.3, 122.2, 122.1, 121.3, 120.8, 118.69, 118.66, 118.62, 118.52, 118.0, 114.1, 108.4, 104.9, 104.8, 104.5, 77.8, 65.21, 65.16, 50.9, 48.83, 48.78, 43.6, 42.2, 38.4, 38.2, 36.2, 36.10, 36.07, 36.0, 34.9, 28.2; HRMS (TOF-ESI<sup>+</sup>) calc'd for C<sub>76</sub>H<sub>87</sub>N<sub>22</sub>O<sub>17</sub> [M+H]<sup>+</sup> 1579.6620, found 1579.6580.

#### 3.4.9 BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-CO<sub>2</sub>H (**13**)

A solution of BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-CO<sub>2</sub>Me **12** (25 mg, 0.0158 mmol) dissolved in 1,4-dioxane (376 μL) and aqueous NaOH (1 N, 253 μL, 0.253 mmol) was stirred at 40 °C for 4 h. The solution was then added to distilled H<sub>2</sub>O (5 mL) preacidified with aqueous HCl (1 N, 253 μL, 0.253 mmol), yielding a precipitate that was diluted with another 15 mL of H<sub>2</sub>O and was then isolated by centrifugation (~4500 rpm). The resultant solid, which contained a small amount of residual H<sub>2</sub>O, was frozen and lyophilized to dryness and then suspended in excess anhydrous Et<sub>2</sub>O, triturated, and filtered, and the filter cake washed with copious amounts of anhydrous Et<sub>2</sub>O. Drying of the tan solid *in vacuo* yielded BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-CO<sub>2</sub>H **13** (23 mg, 93%). <sup>1</sup>H NMR: δ 12.13 (br s, 1H), 10.23 (s, 1H), 10.20 (s, 1H), 9.98-9.90 (m, 6H), 8.01 (m, 1H), 7.443 (s, 1H), 7.439 (s, 1H), 7.42 (d, *J* = 1.7 Hz, 1H), 7.30-7.15 (m, 18H), 7.07 (m, 3H), 6.92 (m, 1H), 6.84 (d, *J* = 2.0 Hz, 1H), 6.81 (t, *J* = 5.6 Hz, 1H), 5.00 (m, 2H), 4.98 (s, 2H), 4.10 (m, 1H), 3.95 (m, 7H), 3.85-3.83 (m, 12H), 3.81 (s, 3H), 3.78 (s, 3H), ~3.30 (m, 2H, obstructed by H<sub>2</sub>O peak), 3.03 (m, 2H), 2.53 (m, 2H), 2.46 (m, 2H), 1.36 (s, 9H); <sup>13</sup>C NMR: δ 167.94, 167.87, 162.4, 162.0, 161.6, 158.54, 158.50, 158.43, 155.8, 155.75, 155.74, 155.6, 155.5, 137.1, 136.0, 133.91, 133.90, 128.3, 127.7, 127.60, 127.57, 123.10, 123.07, 122.8, 122.7, 122.6, 122.3, 122.24, 122.17, 121.16, 121.15, 120.3, 119.5, 118.64, 118.61, 118.5, 118.0, 114.10,

114.06, 108.5, 105.0, 104.5, 77.8, 65.2, 65.1, 48.8, 43.6, 42.2, 38.4, 38.2, 36.2, 36.14, 36.10, 36.08, 36.0, 35.8, 34.9, 28.2; HRMS (TOF-ESI<sup>+</sup>) calc'd for C<sub>75</sub>H<sub>86</sub>N<sub>22</sub>O<sub>17</sub> [M+2H]<sup>2+</sup>/2 783.3271, found 783.3237.

#### 3.4.10 BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-CO<sub>2</sub>Pfp (**14**)

A solution of BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-CO<sub>2</sub>H **13** (250 mg, 0.160 mmol) and DCC (66 mg, 0.320 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (8.8 mL) was stirred at 23 °C for 45 min. The solution was then treated with DMAP (2 mg, 0.016 mmol) followed by pentafluorophenol (175 mg, 0.950 mmol) and stirred at 23 °C for 12 h. The reaction mixture was then loaded onto a silica gel column with CH<sub>2</sub>Cl<sub>2</sub> and eluted with step gradients of 100% CH<sub>2</sub>Cl<sub>2</sub> to 100% acetone with incremental steps of 5% acetone. The product was concentrated *in vacuo* to yield BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-CO<sub>2</sub>Pfp **14** as a brown solid (221 mg, 80%). <sup>1</sup>H NMR: δ 10.20 (s, 1H), 10.16 (s, 1H), 10.08 (s, 1H), 9.99-9.91 (m, 5H), 7.99 (m, 1H), 7.73 (d, *J* = 1.7 Hz, 1H), 7.444 (s, 1H), 7.440 (s, 1H), 7.30-7.12 (m, 20 H), 7.06 (d, *J* = 1.2 Hz, 1H), 7.03 (d, *J* = 8.5 Hz, 1H), 6.92 (s, 1H), 6.80 (t, *J* = 5.4 Hz, 1H), 5.00 (m, 2H), 4.98 (m, 2H), 4.11 (m, 1H), 3.95 (m, 7H), 3.88 (s, 3H), 3.86-3.84 (m, 12H), 3.78 (s, 3H), ~3.30 (m, 2H, obstructed by H<sub>2</sub>O peak), 3.04 (m, 2H), 2.52 (m, 2H), 2.46 (m, 2H), 1.36 (s, 9H); HRMS (TOF-ESI<sup>+</sup>) calc'd for C<sub>81</sub>H<sub>85</sub>F<sub>5</sub>N<sub>22</sub>O<sub>17</sub> [M+2H]<sup>2+</sup>/2 866.3192, found 866.3236.

#### 3.4.11 cyclo-(-ImPyPyPy-(R)<sup>β</sup>-H<sub>2</sub>N<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-H<sub>2</sub>N<sub>γ</sub>-) (**1**)

A solution of BocHN-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-CO<sub>2</sub>Pfp **14** (84 mg, 0.049 mmol) in anhydrous CF<sub>3</sub>CO<sub>2</sub>H:CH<sub>2</sub>Cl<sub>2</sub> (1:1, 4 mL) was stirred at 23 °C for 10 min prior to being concentrated to dryness in a 500 mL round-bottom flask. The residue was then dissolved in cold (0 °C) DMF (10 mL), followed by immediate dilution with MeCN (300 mL) and DIEA (1.6 mL). The reaction mixture was left at 23 °C for 3 days without stirring. [Note: The solution turns cloudy as the macrocyclization proceeds.] The reaction mixture was concentrated to a volume of 11 mL and added to a solution of H<sub>2</sub>O (30 mL) and aqueous HCl (1 N, 9.2 mL) at 0 °C. The protected intermediate cyclo-(-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-) **15** was isolated by centrifugation (~4500 rpm), lyophilized to dryness, and then suspended in excess anhydrous Et<sub>2</sub>O, triturated, and filtered, and the filter cake washed with copious amounts of anhydrous Et<sub>2</sub>O. Drying of the tan solid *in vacuo* yielded the protected intermediate cyclo-(-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-ImPyPyPy-(R)<sup>β</sup>-CbzHN<sub>γ</sub>-) **15**, HRMS (TOF-ESI<sup>+</sup>) calc'd for C<sub>70</sub>H<sub>76</sub>N<sub>22</sub>O<sub>14</sub> [M+2H]<sup>2+</sup>/2 724.2956, found 724.2925. This material

was immediately deprotected by dissolving in  $\text{CF}_3\text{CO}_2\text{H}$  (2 mL) followed by addition of  $\text{CF}_3\text{SO}_3\text{H}$  (100  $\mu\text{L}$ ) at 23 °C for 5 min. The solution was then frozen, and DMF (2 mL) was layered over the frozen solution. The thawed solution was diluted with  $\text{H}_2\text{O}$  (6 mL) and purified by reverse-phase HPLC to give a white solid after lyophilization. The solid was suspended in excess anhydrous  $\text{Et}_2\text{O}$ , triturated, filtered, and the filter cake washed with copious amounts of anhydrous  $\text{Et}_2\text{O}$ . Drying of the white solid *in vacuo* yielded cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -) **1** (46 mg, 68%).  $^1\text{H NMR}$ :  $\delta$  10.56 (s, 2H), 9.91 (s, 4H), 9.88 (s, 2H), 8.17 (t,  $J = 5.6$  Hz, 2H), 7.96 (m, 6H), 7.40 (s, 2H), 7.31 (d,  $J = 1.6$  Hz, 2H), 7.27 (d,  $J = 1.6$  Hz, 2H), 7.19 (d,  $J = 1.6$  Hz, 2H), 7.00 (d,  $J = 1.7$  Hz, 2H), 6.96 (d,  $J = 1.6$  Hz, 2H), 6.94 (d,  $J = 1.7$  Hz, 2H), 3.94 (s, 6H), 3.83 (s, 12H), 3.80 (s, 6H), 3.71 – 3.66 (m, 2H), 3.49 – 3.27 (m, 4H, partially obstructed by  $\text{H}_2\text{O}$  peak), 2.79 (dd,  $J = 16.1$  Hz, 6.0 Hz, 2H), 2.60 (dd,  $J = 15.2$  Hz, 5.2 Hz, 2H). HRMS (TOF-ESI $^+$ ) calc'd for  $\text{C}_{54}\text{H}_{63}\text{N}_{22}\text{O}_{10}$  [M+H] $^+$  1179.5098, found 1179.5087.

*3.4.12 cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -) (3) and cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -) (2)*

A solution of cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -) **1** (2.81 mg, 2.0  $\mu\text{mol}$ ) in anhydrous NMP (200  $\mu\text{L}$ ) and DIEA (20  $\mu\text{L}$ ) at 23 °C was treated with a solution of  $\text{Ac}_2\text{O}$  in NMP (0.122 M, 6.8  $\mu\text{L}$ ). After 10 min, the reaction mixture was treated with another 6.8  $\mu\text{L}$  of  $\text{Ac}_2\text{O}$  in NMP (0.122 M) and allowed to stand for 5 h. The reaction mixture was then diluted to a volume of 10 mL by addition of a 4:1 solution of aqueous  $\text{CF}_3\text{CO}_2\text{H}$  (0.1% v/v):MeCN (5 mL), followed by additional aqueous  $\text{CF}_3\text{CO}_2\text{H}$  (0.1% v/v, 5 mL), and then purified by reverse-phase HPLC to yield cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -) **1** (363 nmol, 18%), cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -) **3** (800 nmol, 40%), and cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -) **2** (432 nmol, 22%). Cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -) **3** HRMS (TOF-ESI $^+$ ) calc'd for  $\text{C}_{56}\text{H}_{65}\text{N}_{22}\text{O}_{11}$  [M+H] $^+$  1221.5203, found 1221.5204. Cyclo-(-ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -) **2** HRMS (TOF-ESI $^+$ ) calc'd for  $\text{C}_{58}\text{H}_{68}\text{N}_{22}\text{O}_{12}$  [M+2H] $^{2+}/2$  633.2646, found 633.2631.

*3.4.13 ImPyPyPy-(*R*) $^{\beta\text{-H}_2\text{N}\gamma}$ -ImPyPyPy-(+)-IPA (4)*

Prepared as described in Chapter 2 of this thesis.<sup>11</sup>

*3.4.14 ImPyPyPy-(*R*) $^{\beta\text{-AcHN}\gamma}$ -ImPyPyPy-(+)-IPA (5)*

A solution of polyamide **4**<sup>11</sup> (7.4 mg, 5.03  $\mu$ moles, assumes **4** as the mono-CF<sub>3</sub>CO<sub>2</sub>H salt) in DMF (1.76 mL) was treated with a solution of Ac<sub>2</sub>O in pyridine (10% v/v, 240  $\mu$ L, 0.254 mmoles Ac<sub>2</sub>O). The solution was allowed to stand at 23 °C for 30 min, and then acidified with aqueous CF<sub>3</sub>CO<sub>2</sub>H (15% v/v, 2 mL). After 5 min the solution was further diluted with distilled H<sub>2</sub>O (5 mL), purified by preparative RP-HPLC, and lyophilized to dryness. Suspension of the residual solid in anhydrous Et<sub>2</sub>O, following by filtration and drying under high vacuum yielded **5** (6.7 mg, 95%). HRMS (FAB<sup>+</sup>) calc'd for C<sub>67</sub>H<sub>79</sub>N<sub>22</sub>O<sub>13</sub> [M+H]<sup>+</sup> 1399.6191, found 1399.6181.

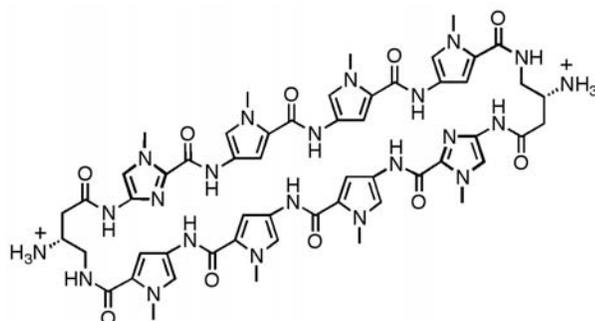
### 3.5 Notes and References

1. Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215–2235.
2. Dervan, P. B., and Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–299.
3. (a) Trauger, J. W., Baird, E. E., and Dervan, P. B. *Nature* **1996**, *382*, 559–561. (b) White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. *Nature* **1998**, *391*, 468–470. (c) Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104–109. (d) Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. *Science* **1998**, *282*, 111–115.
4. Hsu, C. F., Phillips, J. W., Trauger, J. W., Farkas, M. E., Belitsky, J. M., Heckel, A., Olenyuk, B. Z., Puckett, J. W., Wang, C. C. C., and Dervan, P. B. *Tetrahedron* **2007**, *63*, 6146–6151.
5. (a) Belitsky, J. M., Leslie, S. J., Arora, P. S., Beerman, T. A., and Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 3313–3318. (b) Crowley, K. S., Phillion, D. P., Woodard, S. S., Scheitzer, B. A., Singh, M., Shabany, H., Burnette, B., Hippenmeyer, P., Heitmeier, M., and Bashkin, J. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1565–1570. (c) Best, T. P., Edelson, B. S., Nickols, N. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12063–12068. (d) Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., and Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802–2818. (e) Xiao, X., Yu, P., Lim, H. S., Sikder, D., and Kodadek, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865–2868. (f) Nickols, N. G., Jacobs, C. S., Farkas, M. E., and Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 363–370. (g) Dose, C., Farkas, M. E., Chenoweth, D. M., and Dervan, P. B. *J. Am. Chem. Soc.* **2008**, *130*, 6859–6866. (h) Hsu, C. F., and Dervan, P. B. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5851–5855.
6. (a) Gottesfeld, J. M., Melander, C., Suto, R. K., Raviol, H., Luger, K., and Dervan, P. B. Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides. *J. Mol. Biol.* **2001**, *309*, 615–629. (b) Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* **2003**, *326*, 371–380. (c) Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6864–6869. (d) Dudouet, B., Burnett, R., Dickinson, L. A., Wood, M. R., Melander, C., Belitsky, J. M., Edelson, B., Wurtz, N., Briehn, C., Dervan, P. B., and Gottesfeld, J. M. *Chem. Biol.* **2003**, *10*, 859–867.
7. (a) Olenyuk, B. Z., Zhang, G. J., Klco, J. M., Nickols, N. G., Kaelin, Jr., W. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16768–16773. (b) Kageyama, Y., Sugiyama, H., Ayame, H., Iwai, A., Fujii, Y., Huang, L. E., Kizaka-Kondoh, S., Hiraoka, M., and Kihara, K. *Acta. Oncol.* **2006**, *45*, 317–324. (c) Nickols, N. G., Jacobs, C. S., Farkas, M. E., and Dervan, P. B. *ACS Chem.*

*Biol.* **2007**, *2*, 561–571.

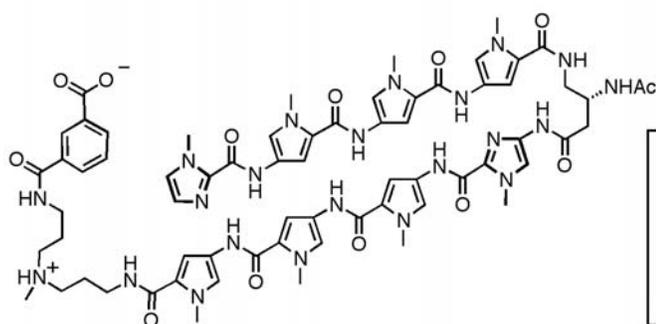
8. Nickols, N. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10418–10423.
9. (a) Matsuda, H., Fukuda, N., Ueno, T., Tahira, Y., Ayame, H., Zhang, W., Bando, T., Sugiyama, H., Saito, S., Matsumoto, K., and others, O. *J. Am. Soc. Nephrol.* **2006**, *17*, 422–432. (b) Yao, E. H., Fukuda, N., Ueno, T., Matsuda, H., Matsumoto, K., Nagase, H., Matsumoto, Y., Takasaka, A., Serie, K., Sugiyama, H., and Sawamura, T. *Hypertension* **2008**, *52*, 86–92.
10. (a) Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10389–10392. (b) Zhang, Q., Dwyer, T. J., Tsui, V., Case, D. A., Cho, J., Dervan, P. B., Wemmer, D. E. *J. Am. Chem. Soc.* **2004**, *126*, 7958–7966. (c) Herman, D. M., Turner, J. M., Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121–1129. (d) Melander, C., Herman, D. M., Dervan, P. B. *Chem. Eur. J.* **2000**, *6*, 4487–4497.
11. Chenoweth D. M., Harki, D. A., and Dervan, P. B. Solution-phase synthesis of pyrrole-imidazole polyamides. *In Press*. (See Chapter 2 of this thesis)
12. (a) Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141–6146. (b) Wurtz, N. R., Turner, J. M., Baird, E. E., Dervan, P. B. *Org. Lett.* **2001**, *3*, 1201–1203. (c) Jaramillo, D., Liu, Q., Aldrich-Wright, J., Tor, Y. *J. Org. Chem.* **2004**, *69*, 8151–8153.
13. Trauger, J. W., Dervan P. B. *Methods Enzymol.* **2001**, *340*, 450–466.
14. For examples of polyamides with  $K_a$  values  $> 2 \times 10^{10} \text{ M}^{-1}$  and a discussion of the limitations of quantitative DNase I footprint titrations, please refer to reference 5g. An analogous polyamide to **4**, ImPyPyPy-(R) $^{\beta\text{-H}_2\text{N}\gamma}$ -ImPyPyPy- $\beta$ -Dp, was found to have  $K_a > 2 \times 10^{10} \text{ M}^{-1}$ .<sup>5g</sup> Additionally, previous studies with cyclic polyamide cyclo(-ImPyPyPy-(R) $^{\alpha\text{-H}_2\text{N}\gamma}$ -ImPyPyPy- $\gamma$ -) found  $K_a$  values far exceeding  $2 \times 10^{10} \text{ M}^{-1}$  by DNase I footprint titrations.<sup>10c,d</sup> Cyclic polyamide **1** possesses dual  $\beta$ -amino functionalities; a modification that yields even greater DNA binding affinities compared with  $\alpha$ -amino and unsubstituted  $\gamma$ -turns for hairpin polyamides of sequence ImPyPyPy- $\gamma$ -ImPyPyPy.<sup>5g</sup> The DNA binding affinity of **1** most likely supersedes that of predecessor cyclo(-ImPyPyPy-(R) $^{\alpha\text{-H}_2\text{N}\gamma}$ -ImPyPyPy- $\gamma$ -).
15. (a) Pilch, D. S., Poklar, N., Gelfand, C. A., Law, S. M., Breslauer, K. J., Baird, E. E., Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306–8311. (b) Pilch, D. S., Poklar, N., Baird, E. E., Dervan, P. B., Breslauer, K. J. *Biochemistry* **1999**, *38*, 2143–2151.
16. Shaffer, P. L., Jivan, A., Dollins, D. E., Claessens, F., Gewirth, D. T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4758–4763.
17. For a review of pharmacokinetics in drug discovery see: Ruiz-Garcia, A., Bermejo, M., Moss, A., Casabo, V.G. *J. Pharm. Sci.* **2008**, *97*, 654–690.
18. Apredica, 313 Pleasant St., Watertown, MA 02472 (<http://www.apredica.com/>).
19. Harki, D. A., Satyamurthy, N., Stout, D. B., Phelps, M. E., Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13039–13044.

### 3.6 Spectra and Supplemental Information



*Cyclic polyamide 1*

Denoted as **DMC2-239** in ADMET data tables (Table 3.2-3.7) and in full ADMET report located in Appendix B



*Hairpin polyamide 5*

Denoted as **DH-V-88** in ADMET data tables (Table 3.2-3.7) and in full ADMET report located in Appendix B

**Figure 3.3** Polyamides **1** and **5** were subjected to preclinical ADMET testing by contract service at Apredica (Watertown, MA). Shown below (Table 3.2-3.7) are summaries of the ADMET results taken directly from the final report provided by Apredica. The full ADMET report, which includes experimental conditions, can be found in Appendix B of this thesis.

**Table 3.2** Caco-2 permeability summary from Apredica report (Appendix B).

Client ID	test conc (μM)	Assay duration (hr)	mean A->B $P_{app}^a$ ( $10^{-6}$ cm s $^{-1}$ )	mean A->B $P_{app}^a$ ( $10^{-6}$ cm s $^{-1}$ )	Asymmetry ratio <sup>b</sup>	comment
Warfarin	50	2	35.4	7.9	0.2	high permeability control
Ranitidine	50	2	1.4	2.4	1.7	low permeability control
DH-V-88	10	2	ND	0.11	UD	
DMC2-239	10	2	ND	ND	ND	

<sup>a</sup>Apparent permeability

<sup>b</sup> $P_{app}(B \rightarrow A) / P_{app}(A \rightarrow B)$

ND = no compound detected in receiver solution

**Table 3.3** Cytotoxicity summary from Apredica report (Appendix B).

Client ID	Cell line	IC50 (μM)	comment
Chlorpromazine	HepG2	13	Higly cytotoxic control
Propranolol	HepG2	80	Low cytotoxic control
DH-V-88	HepG2	>100	
DMC2-239	HepG2	>100	

**Table 3.4** Fluorescent Cyp IC50 summary from Apre dica report (Appendix B).

Client ID	IC <sub>50</sub> (μM)						
	Cyp1A2 / CEC	Cyp2C8/D BF	Cyp2C9 / DBF	Cyp2C19 / DBF	Cyp2D6 / AMMC	Cyp3A4 / BFC	Cyp3A4 / DBF
Controls	0.2 α-naphtho- flavone	2.3 ketoconazol e	1.1 sulpha- phenazole	5.6 tranyl- cypromine	0.05 quinindine	1.26 ketoconazol e	1.26 ketoconazol e
DH-V-88	>50	>50	>50	>50	>50	47.6	>50
DMC2-239	>50	>50	>50	>50	>50	37.7	>50

**Table 3.5** hERG FastPatch summary from Apre dica report (Appendix B).

Client ID	IC <sub>50</sub> (μM)	comment
	99% at 0.5 μM	
E-4031		positive control
DH-V-88	>100	*
DMC2-239	>100	*

\*The solubility limit for this experiment, as determined by vehicle controls, was 17.3 x 10<sup>3</sup> LSU (horizontal black line). Based on the data obtained, there may be solubility issues for both test articles at 30 and 100 μM in our physiological saline solution (HB-PS, 0.3%DMSO). Precipitation of DH-V-88 at 100 μM was visible to the naked eye.

**Table 3.6** Plasma half-life summary from Apre dica report (Appendix B).

Compound	test conc (μM)	medium	T1/2 (min)	Fraction remaining, max time (%)	comment
Propantheline	10.0	Human Plasma	35.5	5.8%	control
Propantheline	10.0	Rat Plasma	149.0	51.6%	control
DH-V-88	10.0	Human Plasma	>120	95.6%	
DH-V-88	10.0	Rat Plasma	>120	94.0%	
DMC2-239	10.0	Human Plasma	>120	124.5%	
DMC2-239	10.0	Rat Plasma	>120	120.3%	

<sup>a</sup>Half-life

**Table 3.7** Plasma protein binding summary from Apre dica report (Appendix B).

Client ID	test conc (μM)	Assay duration	Species	Mean free fraction (%)	comment
Warfarin	10	4 hr	Human	0.73%	high binding control
Warfarin	10	4 hr	Rat	5.47%	high binding control
Atenolol	10	4 hr	Human	76.2%	low binding control
Atenolol	10	4 hr	Rat	84.7%	low binding control
DH-V-88	10	4 hr	Human	0.0015%	
DH-V-88	10	4 hr	Rat	0.0016%	
DMC2-239	10	4 hr	Human	0.0000%	
DMC2-239	10	4 hr	Rat	0.0040%	

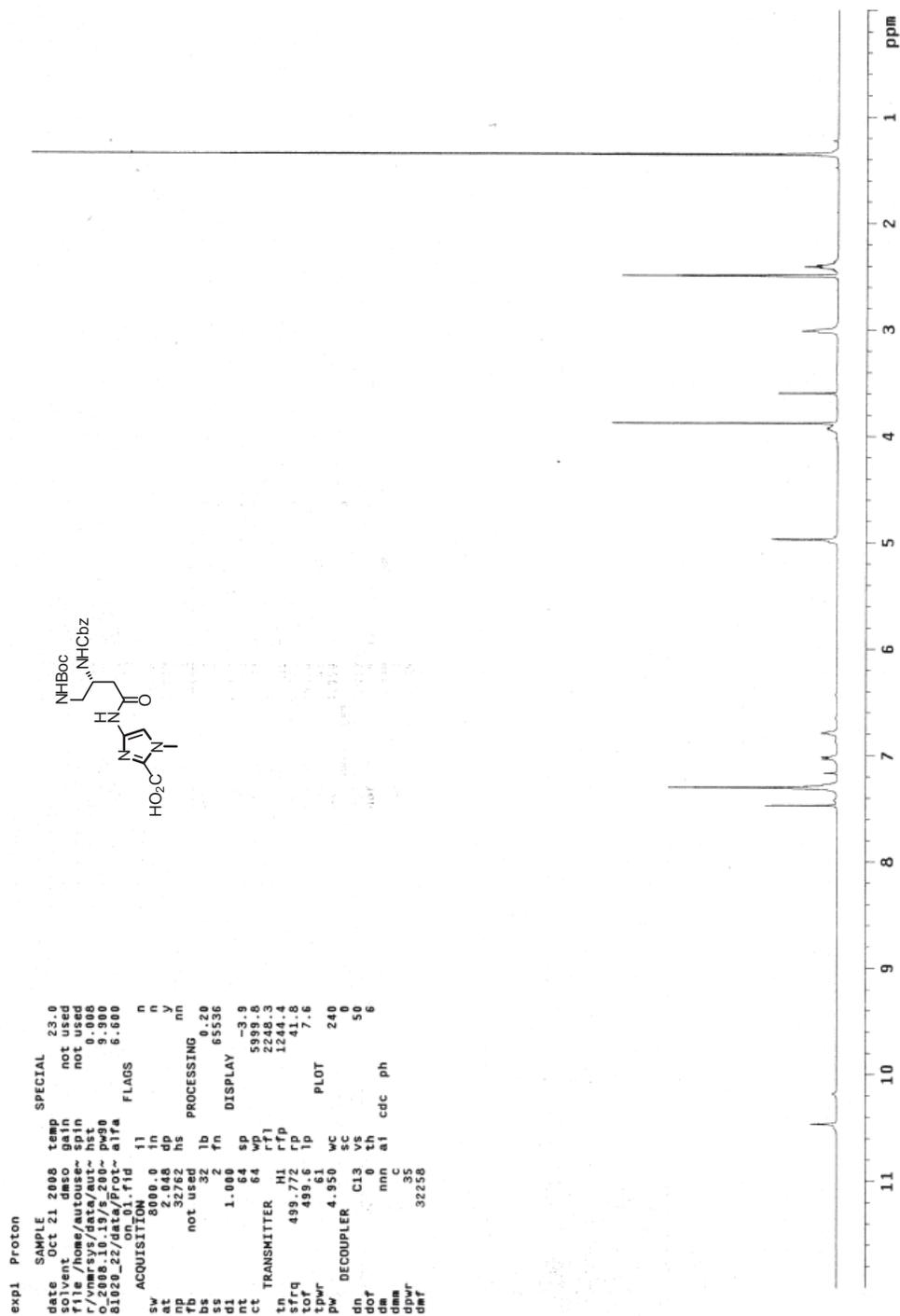
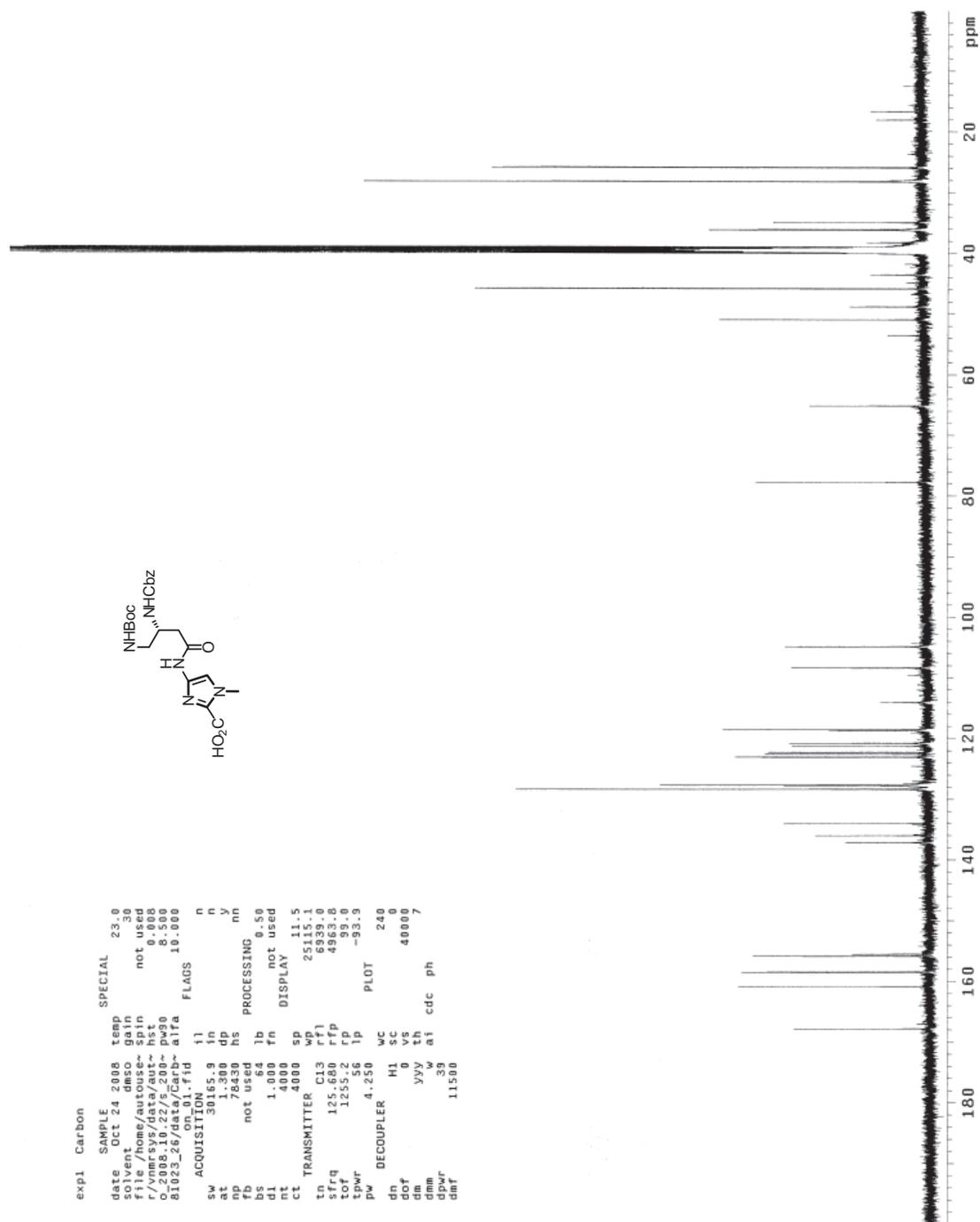
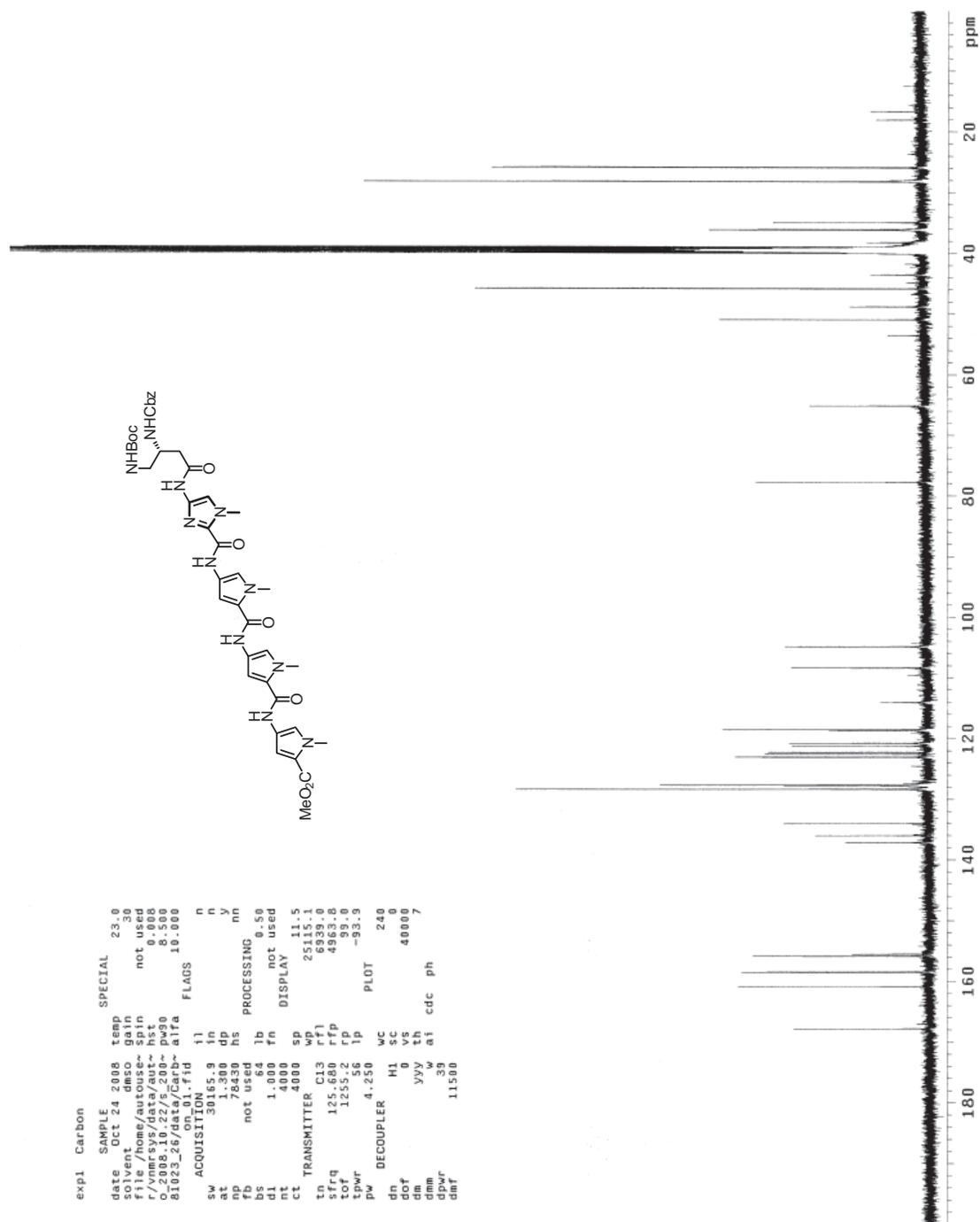


Figure 3.4  $^1\text{H}$  NMR BocHN-(*R*) $^{\beta}$ -CbzHN $^{\gamma}$ -Im-CO $_2$ H (8)



**Figure 3.5** <sup>13</sup>C NMR BocHN-(R)<sup>β</sup>-CbzHN-γ-Im-CO<sub>2</sub>H (8)





**Figure 3.7** <sup>13</sup>C NMR BocHN-(R) $\beta$ -CbzHN- $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me (9)

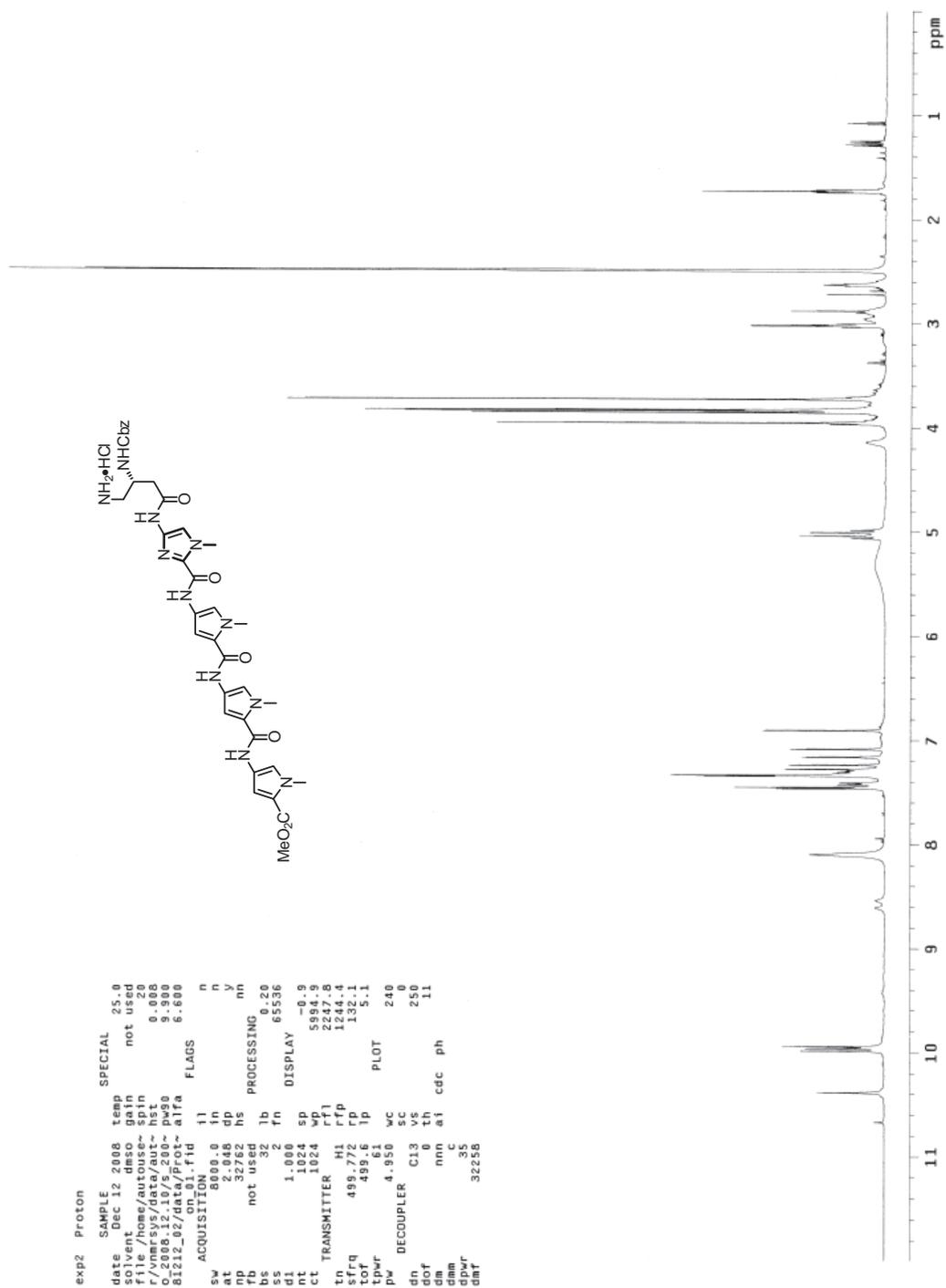
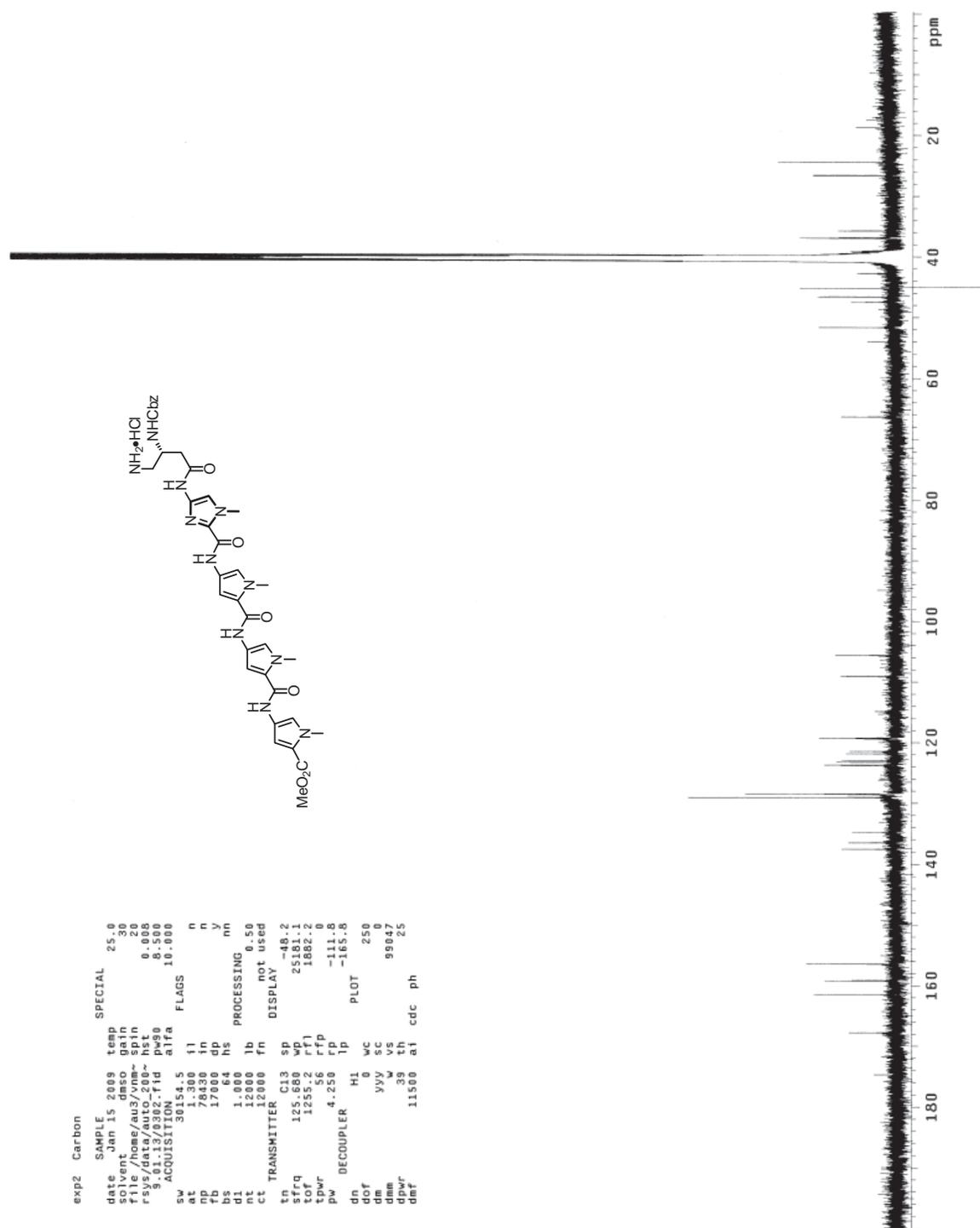


Figure 3.8  $^1\text{H}$  NMR  $\text{HCl}\cdot\text{H}_2\text{N}-(R)^\beta\text{-CbzHN}\gamma\text{-ImPyPyPy-CO}_2\text{Me}$  (10)



**Figure 3.9**  $^{13}\text{C}$  NMR  $\text{HCl}\cdot\text{H}_2\text{N}-(R)\beta\text{-CbzHN}\gamma\text{-ImPyPyPy-CO}_2\text{Me}$  (10)

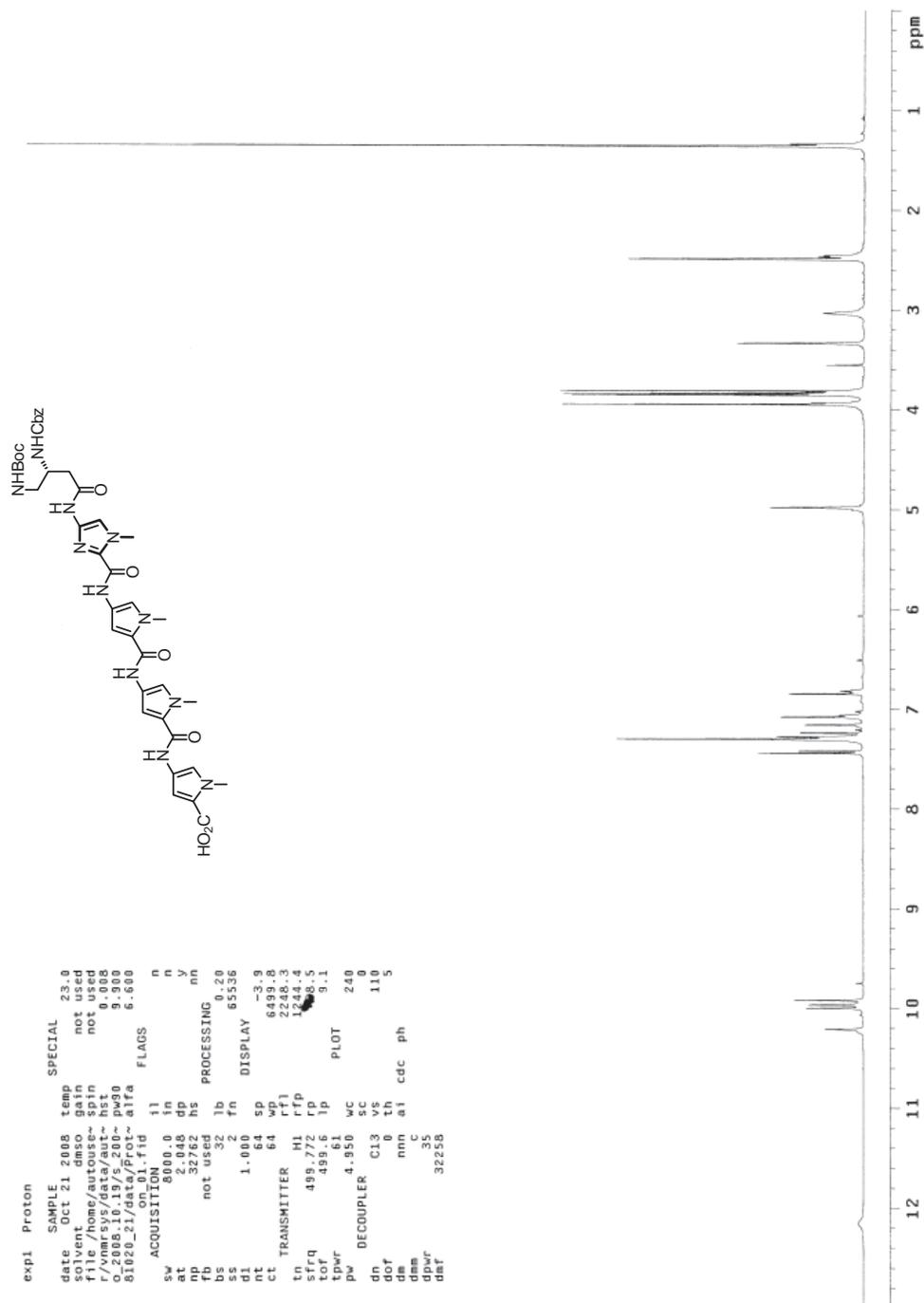
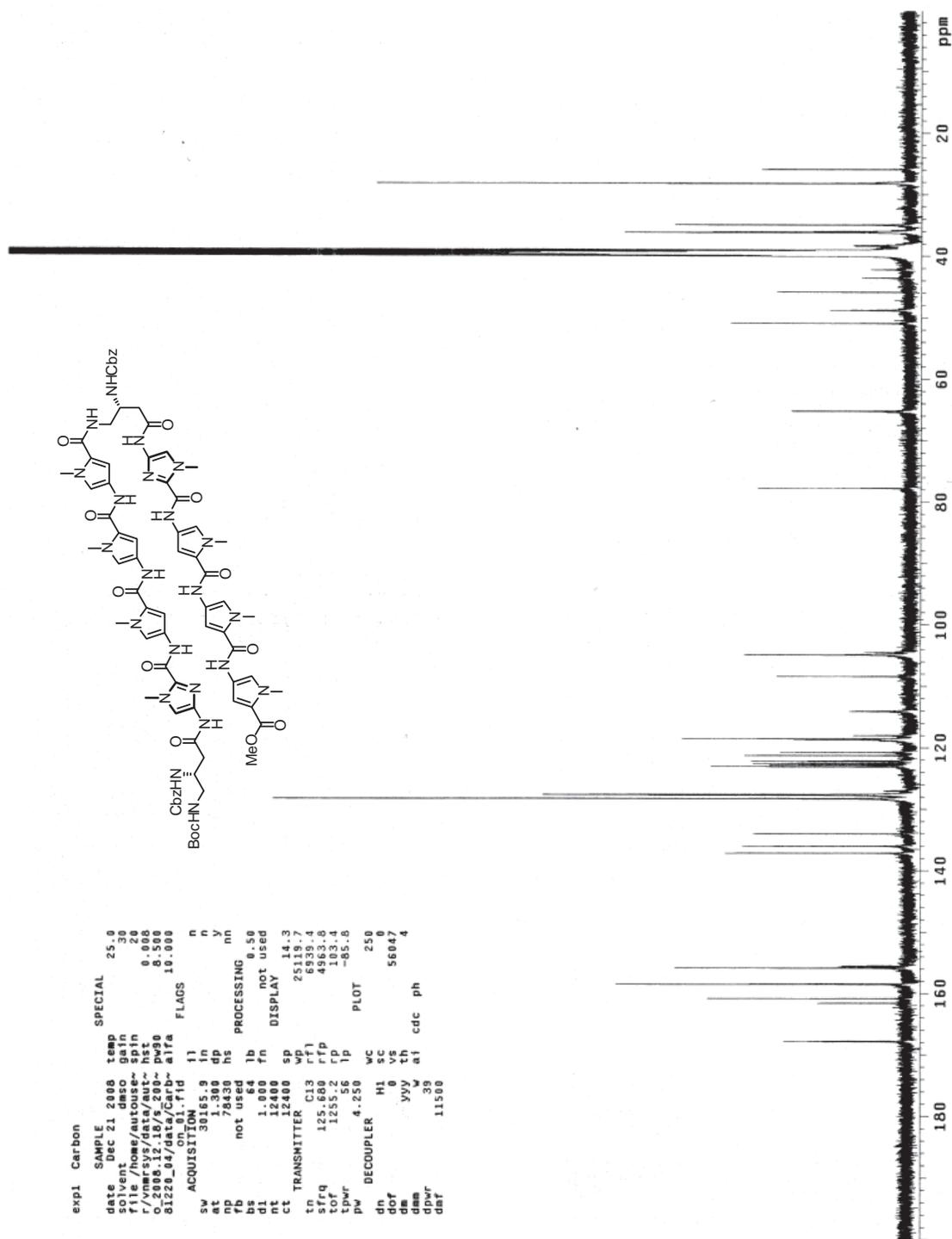


Figure 3.10 <sup>1</sup>H NMR BocHN-(*R*) $\beta$ -CbzHN- $\gamma$ -ImPyPyPy-CO<sub>2</sub>H (11)







**Figure 3.13**  $^{13}\text{C}$  NMR BocHN-(*R*)- $\beta$ -CbzHN- $\gamma$ -ImPyPyPy-(*R*)- $\beta$ -CbzHN- $\gamma$ -ImPyPyPy-CO<sub>2</sub>Me (12)

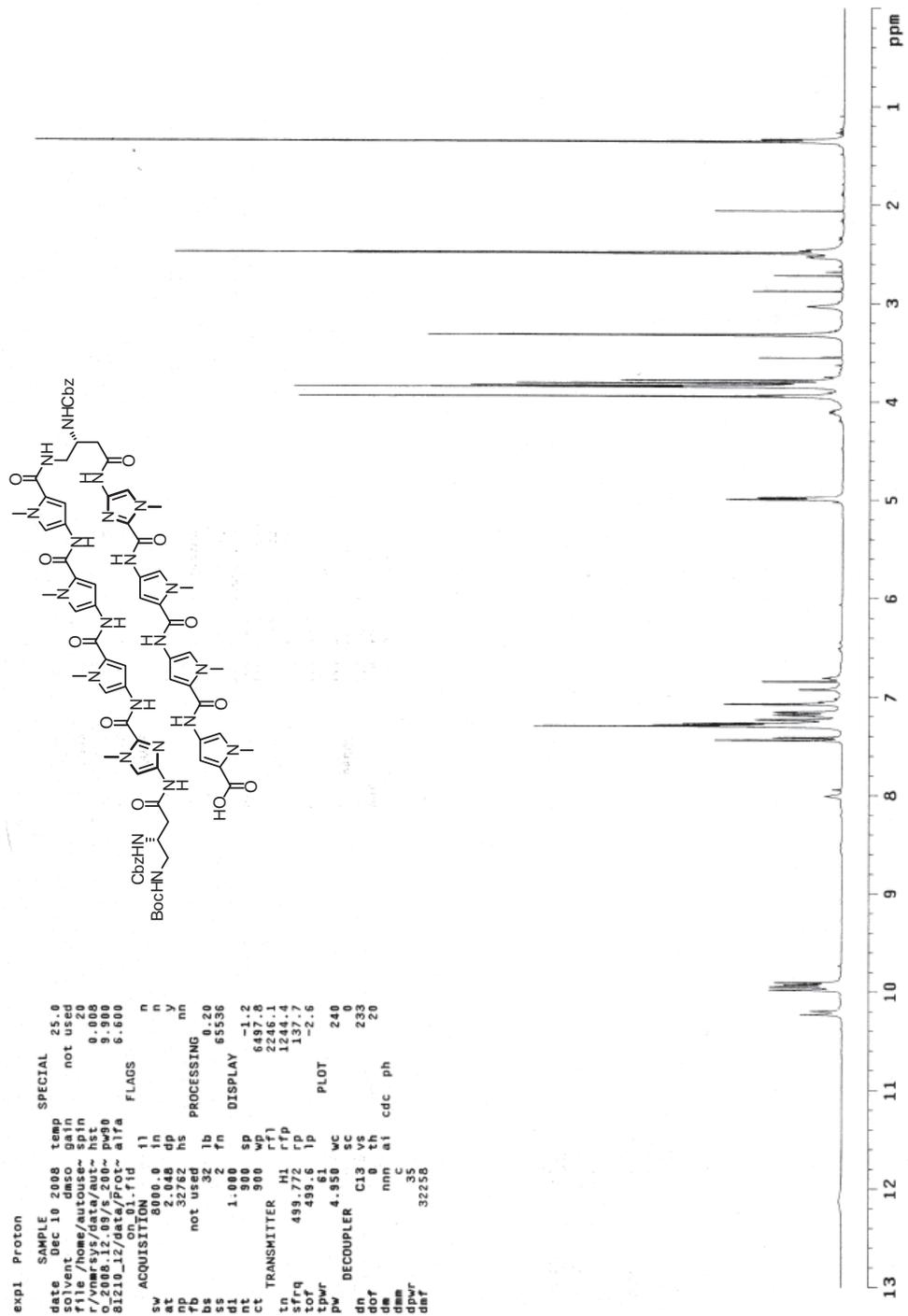


Figure 3.14 <sup>1</sup>H NMR BocHN-(R) $\beta$ -CbzHN- $\gamma$ -ImPyPyPy-(R) $\beta$ -CbzHN- $\gamma$ -ImPyPyPy-CO<sub>2</sub>H (13)



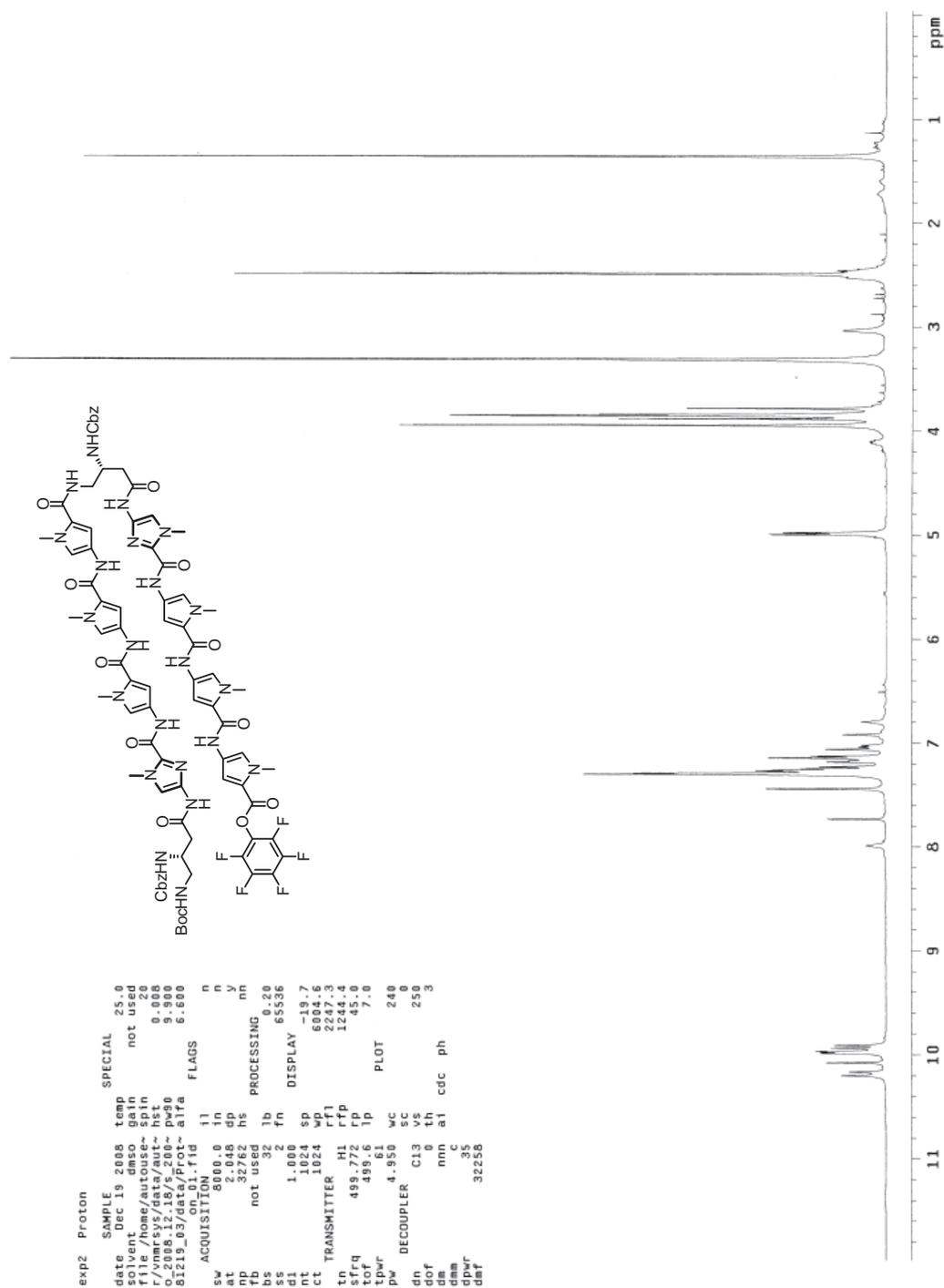
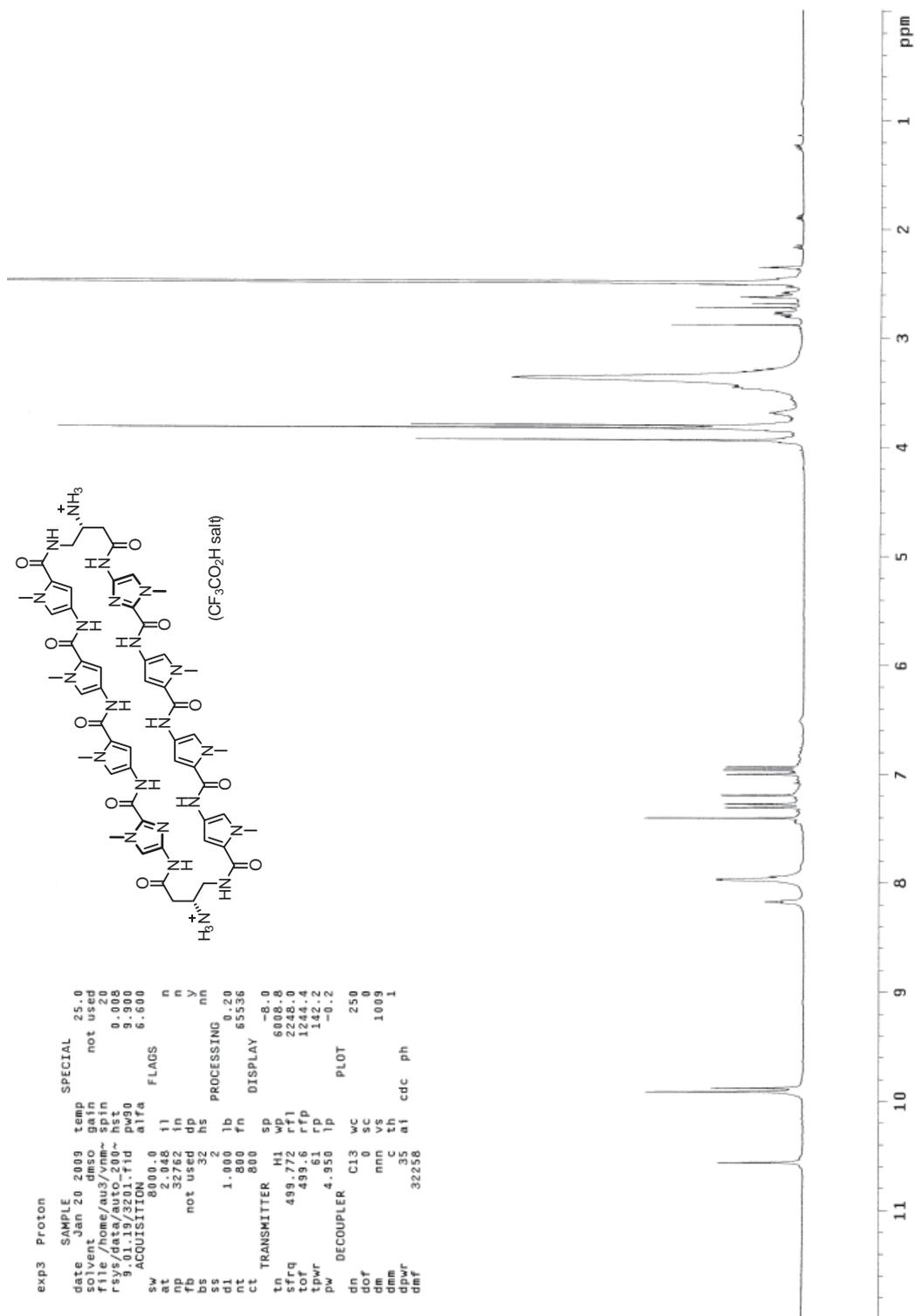
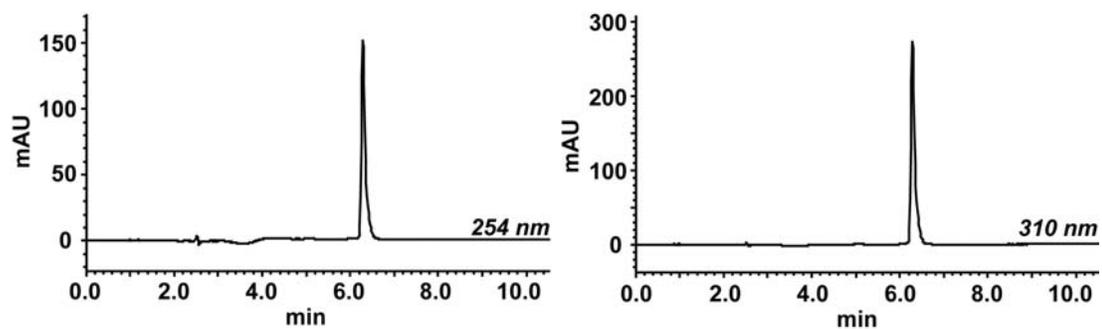


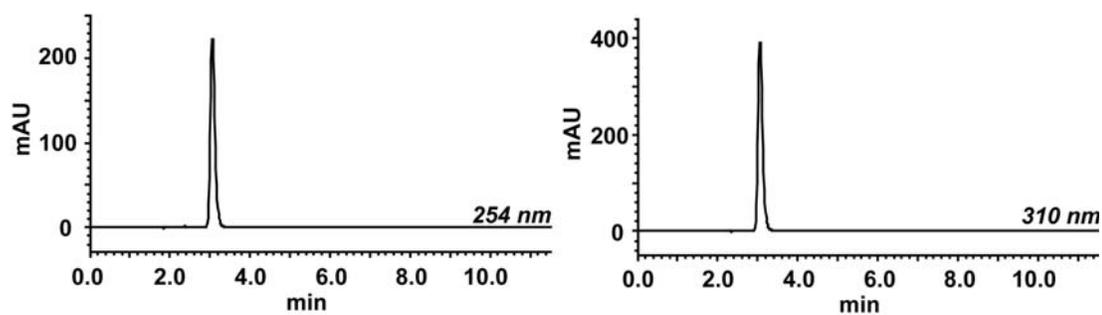
Figure 3.16 <sup>1</sup>H NMR BocHN-(*R*)<sup>β</sup>-CbzHN-γ-ImPyPyPy-(*R*)<sup>β</sup>-CbzHN-γ-ImPyPyPy-CO<sub>2</sub>Pfp (14)



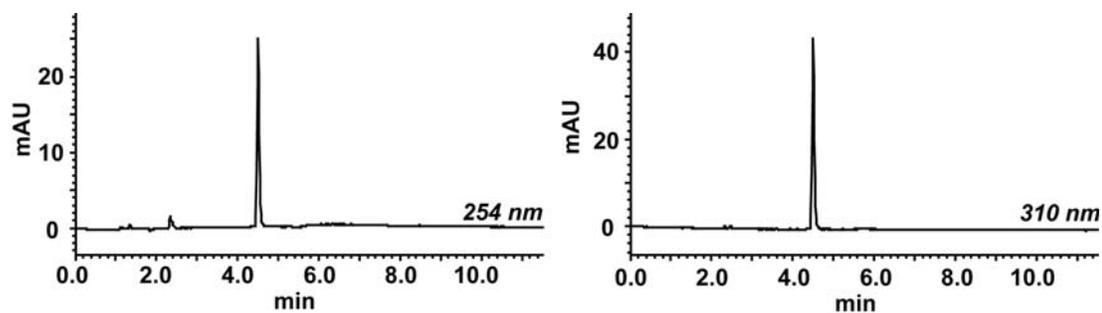
**Figure 3.17** <sup>1</sup>H NMR *cyclo*-(ImPyPyPy-(R)<sup>β</sup>-H<sub>2</sub>N<sup>γ</sup>-ImPyPyPy-(R)<sup>β</sup>-H<sub>2</sub>N<sup>γ</sup>-) (1)



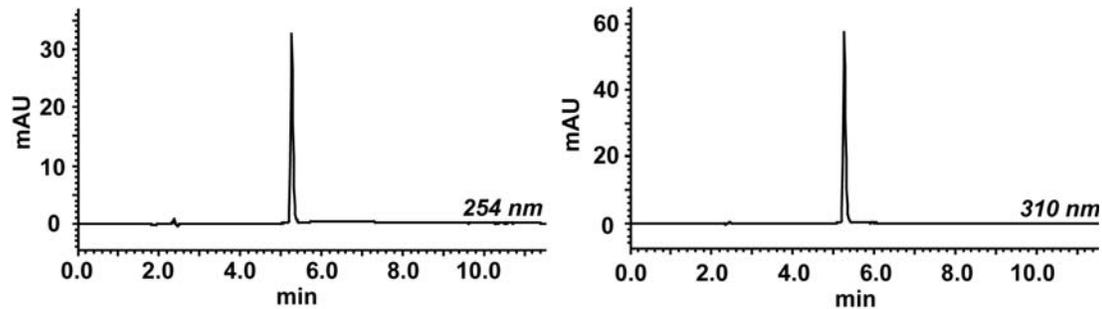
**Figure 3.18** Analytical HPLC characterization of cyclic polyamide 15.



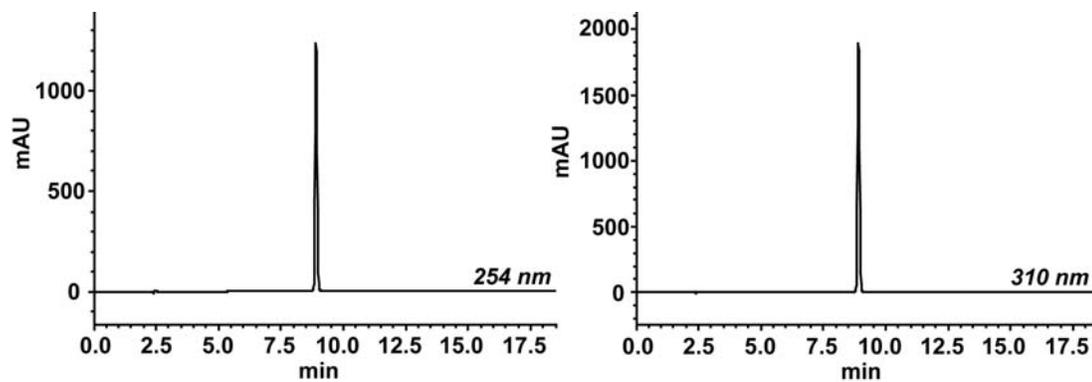
**Figure 3.19** Analytical HPLC characterization of cyclic polyamide 1.



**Figure 3.20** Analytical HPLC characterization of cyclic polyamide 3.



**Figure 3.21** Analytical HPLC characterization of cyclic polyamide 2.



**Figure 3.22** Analytical HPLC characterization of cyclic polyamide 5.

## Chapter 4: Oligomerization Route to Polyamide Macrocycles

*The text of this chapter was taken in part from a manuscript coauthored with Daniel A. Harki and Peter B. Dervan\* (Caltech)*

*(Chenoweth, D.M., Harki, D.A., Dervan, P. B. In Preparation.)*

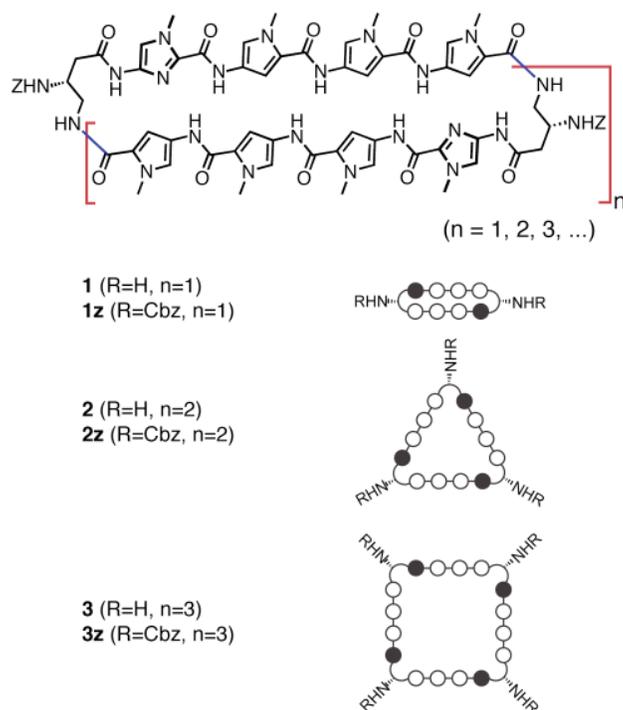
**Abstract**

Cyclic pyrrole–imidazole polyamides are sequence-specific DNA-binding small molecules that are cell permeable and can regulate endogenous gene expression. Syntheses of cyclic polyamides have been achieved by solid-phase and more recently, solution-phase methods. We report a rapid solution-phase oligomerization approach to cyclic polyamides that yields 8, 12, and 16 membered macrocycles. A remarkable preference for DNA binding by the 8 and 16 membered oligomers was observed over the 12-ring macrocycle, which we attributed to a conformational constraint not present in the smaller and larger systems.

## 4.1 Introduction

Pyrrole–imidazole polyamides are a class of cell-permeable small molecules that bind to the minor groove of DNA in a sequence specific manner.<sup>1,2</sup> Antiparallel arrangements of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamides (Im/Py) recognize G•C from C•G base pairs, whereas Py/Py specifies for both T•A and A•T.<sup>3</sup> Hairpin Py-Im polyamides have been programmed for a broad repertoire of DNA sequences with high affinities.<sup>4</sup> These cell permeable<sup>5</sup> ligands can influence gene transcription by disrupting protein-DNA interfaces,<sup>2</sup> and have been shown to control transcription of genes important in human disease.<sup>6</sup> Py-Im polyamides have also been used for a variety of applications ranging from DNA detection<sup>7</sup> and transcriptional activation<sup>8</sup> to the self-assembly of higher order structures.<sup>9</sup> Conjugation of polyamides to functional domains have yielded artificial transcription factor mimics,<sup>8</sup> fluorescent sequence-specific DNA probes,<sup>7</sup> and DNA nano-architectures.<sup>9</sup>

We recently reported solution-phase methods for the synthesis of hairpin<sup>10</sup> and cyclic polyamides.<sup>11</sup> Key to the cyclic polyamide synthesis was a highly efficient macrocyclization



**Figure 4.1** Structures of macrocyclic polyamides **1z–3z** and **1–3**, and their ball-and-stick models. Polyamide shorthand code: closed circles, Im; open circles, Py monomers. (Z = benzyl carbamate protecting group)

that yielded polyamide **1z** (Figure 4.1) from an acyclic precursor. Activation of the C-terminal Py amino acid of **1z** as a pentafluorophenyl ester allowed efficient macrocyclization by the  $\gamma$ -NH<sub>2</sub> on the turn moiety under dilute reaction conditions. Our studies of polyamide **1**, the deprotected analogue of **1z**, revealed it possessed extremely high DNA binding affinities, was cell permeable, and could disrupt androgen receptor-activated gene expression in cell culture. Additionally, preliminary studies of the in vitro ADMET properties of **1** revealed excellent metabolic stability.<sup>11,12</sup>

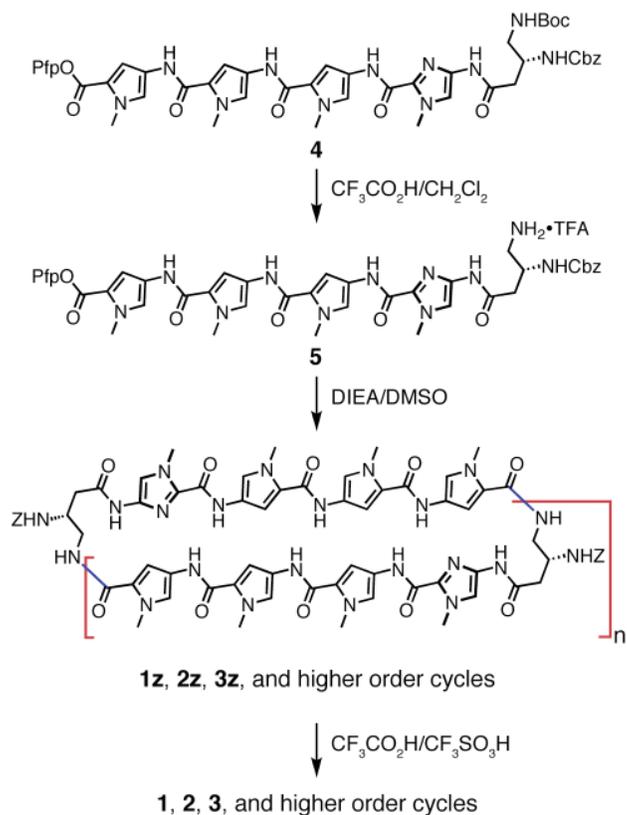
An orthogonal polymerization/oligomerization strategy for the synthesis of **1** and related polyamides is reported here. This method delivers symmetrical Py-

In polyamide macrocycles from simple Py-Im building blocks in a rapid and convergent manner (Figure 4.2). Separable higher-order oligomers, such as the 12-membered (**2**) and 16-membered (**3**) cyclic polyamides are also produced by this method. In addition to describing the synthetic chemistry to prepare **1–3**, we report for the first time the DNA binding properties of such expanded polyamide macrocycles.

## 4.2 Results and Discussion

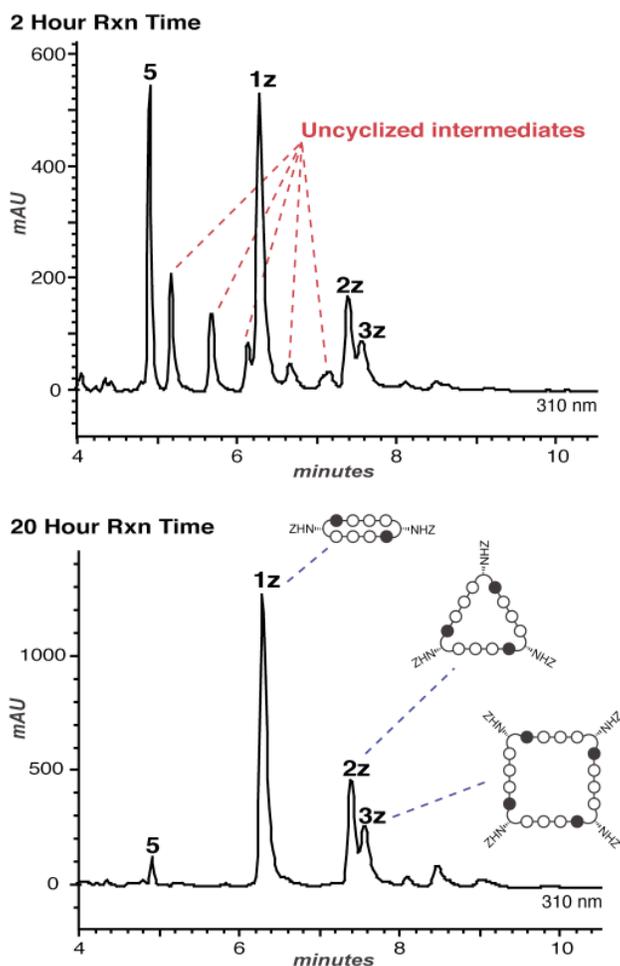
Our strategy for this oligomerization route relied on the palindromic nature of polyamide **1**. Disconnection of **1** at both  $\gamma$ -amino turns affords two identical halves of the molecule. Bimolecular coupling between two molecules, followed by intramolecular ring closure delivers cyclic Py-Im polyamides. Bifunctional monomer **4** contains every atom needed to construct cyclic polyamides **1–3** by this process (Figure 4.2).

The pentafluorophenyl ester **4** was prepared in one step from the previously reported carboxylic acid of **4**.<sup>11</sup> Acidic deprotection of the  $\gamma$ -amino functionality of **4** followed by drying



**Figure 4.2** Synthesis of macrocyclic polyamides **1–3** and higher order cycles by oligomerization of bifunctional intermediate **5**.

*in vacuo* yields intermediate **5** which is the substrate for the homodimerization/oligomerization reaction. To initiate this sequence, the protected trifluoroacetate salt **5** was diluted with DMSO, then treated with an organic base (DIEA) to unmask the highly nucleophilic primary  $\gamma$ -amine. The ensuing oligomerization/macrocyclization process provides benzylcarbamate protected cyclic polyamides **1z**, **2z**, **3z**, and trace amounts of unisolated higher-order oligomers. A distribution of uncyclized intermediates corresponding to the dimer (8-ring cycle, **1z**), trimer (12-ring cycle, **2z**), tetramer (16-ring cycle, **3z**), and higher order adducts can be observed at early time points, as evidenced by HPLC analysis at 2 hr (Figure 4.3). Extended



**Figure 4.3** (Top) Reverse phase HPLC analysis (2 hr time point) of the oligomerization reaction showing **1z**, **2z**, **3z**, and acyclic intermediates. (Bottom) Analysis (20 h time point) of the oligomerization reaction revealing **1z**, **2z**, **3z**, and consumption of uncyclized intermediates. Peaks were identified by high-resolution mass spectrometry following separation and Cbz-deprotection.

reaction times (20 hours) reveals cyclized polyamides **1z**, **2z**, and **3z** in a ratio of 6.6:2.6:1 almost exclusively (Figure 4.3). Isolation of **1z** (13.9%), **2z** (5.5%), and **3z** (2.1%) by preparative HPLC, followed by Cbz-deprotection under acidic conditions ( $\text{CF}_3\text{CO}_2\text{H}/\text{CF}_3\text{SO}_3\text{H}$ ) provides polyamide macrocycles **1–3**.

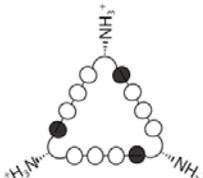
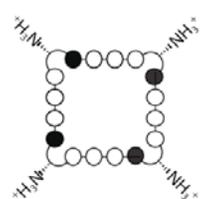
With polyamide macrocycles **2** and **3** in hand we evaluated their ability to bind duplex DNA relative to previously reported cycle **1**.<sup>11</sup> Polyamide **1** increases the dsDNA melting temperature by 23.6 °C. To our surprise, trimeric macrocycle **2** failed to bind its target double stranded DNA sequence as evidenced by the complete lack of ligand-promoted thermal stabilization of duplex DNA melting (Table 4.1). This result is presumably due to inherent geometrical constraints of **2**, preventing the side-by-side antiparallel alignment of the PyPyPyIm strands, a motif well accommodated by the DNA minor groove. In the case of tetrameric macrocycle **3**, dsDNA binding and thermal stabilization was completely restored to a comparable

value to dimer **1**. We hypothesize that an even number of PyPyPyIm strands allows **3** to possess a collapsed or folded tetramer geometry, with two adjacent, antiparallel PyPyPyIm strands followed by an identical repeat of this motif linked through two intervening turn units.

#### 4.3 Conclusion

In summary, we have demonstrated that macrocyclic polyamides can be synthesized by

**Table 4.1**  $T_m$  values for cycles 1–3 in the presence of DNA.<sup>a</sup>

dsDNA sequence =		5' -TTGC TGTTC T GCAA-3'		3' -AACG ACAAGA CGTT-5'
Polyamide cycle		$T_m / ^\circ\text{C}$		$\Delta T_m / ^\circ\text{C}$
—		60.0 ( $\pm 0.3$ )	—	
	(1)	83.5 ( $\pm 0.5$ )	23.6 ( $\pm 0.6$ )	
	(2)	60.1 ( $\pm 0.6$ )	0.1 ( $\pm 0.6$ )	
	(3)	83.0 ( $\pm 0.3$ )	23.1 ( $\pm 0.4$ )	

<sup>a</sup>All values reported are derived from at least three melting temperature experiments with standard deviations indicated in parentheses.  $\Delta T_m$  values are given as  $T_m^{(\text{DNA/polyamide})} - T_m^{(\text{DNA})}$ . The propagated error in  $\Delta T_m$  measurements is the square root of the sum of the square of the standard deviations for the  $T_m$  values.

oligomerization of a bifunctional polyamide to yield a distribution of cyclic polyamide oligomers. Additionally, we show that certain cyclic polyamide geometries are completely devoid of the ability to bind dsDNA, a result which could be utilized in the design of highly specific molecules for targeting non-B-form DNA structures or other higher-order nucleic acid motifs.

## 4.4 Experimental Section

### 4.4.1 General

Chemicals were purchased from Sigma-Aldrich and were used without further purification. (*R*)-3,4-Cbz-Dbu(Boc)-OH was purchased from Senn Chemicals AG (code number 44159). Bulk grade solvents were from Fisher Scientific. Analytical HPLC analysis was conducted on a Beckman Gold instrument equipped with a Phenomenex Gemini analytical column (250 × 4.6 mm, 5 μm), a diode array detector, and the mobile phase consisted of a gradient of acetonitrile (MeCN) in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H. Preparative HPLC was performed on an Agilent 1200 system equipped with a solvent degasser, diode array detector, and a Phenomenex Gemini column (250 × 21.2 mm, 5 μm). A gradient of MeCN in 0.1% (v/v) aqueous CF<sub>3</sub>CO<sub>2</sub>H was utilized as the mobile phase. NMR spectroscopy was performed on a Varian instrument operating at 499.8 MHz (for <sup>1</sup>H) at ambient temperature. All NMR analyses were performed in DMSO-*d*<sub>6</sub>, and chemical shifts are reported in parts per million relative to the internal solvent peak referenced to 2.49 (for <sup>1</sup>H). High-resolution mass spectrometry (HRMS) was recorded in positive-ion mode by fast-atom bombardment (FAB+) on a JEOL JMS-600H instrument or by matrix-assisted, LASER desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) on an Applied Biosystems Voyager DE-Pro spectrometer

using  $\alpha$ -cyano-4-hydroxycinnamic acid as matrix.

#### 4.4.2 BocHN-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>H

Synthesized as previously described in Chapter 3 of this thesis.

#### 4.4.3 BocHN-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Pfp (**4**)

A solution of BocHN-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>H (100 mg, 0.119 mmol) and DCC (49 mg, 0.238 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (5.2 mL) was stirred at 23 °C for 45 min. The solution was then treated with DMAP (1.4 mg, 0.012 mmol) followed by pentafluorophenol (131.2 mg, 0.713 mmol) and stirred at 23 °C for 12 h. The reaction mixture was then loaded onto a silica gel column with CH<sub>2</sub>Cl<sub>2</sub> and eluted with step gradients of 100% CH<sub>2</sub>Cl<sub>2</sub> to 100% acetone with incremental steps of 5% acetone. The product was concentrated *in vacuo* to yield BocHN-(R) <sup>$\beta$</sup> -CbzHN $\gamma$ -ImPyPyPy-CO<sub>2</sub>Pfp (**4**) as an off-white solid (84 mg, 71%). <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  10.16 (s, 1H), 10.08 (s, 1H), 9.99 (s, 1H), 9.97 (s, 1H), 7.74 (d, J = 1.8 Hz, 1H), 7.44 (s, 1H), 7.33 – 7.30 (m, 5H), 7.29 (d, J = 1.9 Hz, 1H), 7.27 (d, J = 1.8 Hz, 1H), 7.25 (d, J = 1.8 Hz, 1H), 7.14 (d, J = 1.8 Hz, 1H), 7.13 (d, J = 1.8 Hz, 1H), 7.03 (d, J = 8.4 Hz, 1H), 6.80 (t, J = 5.8 Hz, 1H), 4.98 (s, 2H), 3.95 (m, 4H), 3.89 (s, 3H), 3.860 (s, 3H), 3.856 (s, 3H), 3.03 (m, 2H), 2.46 (m, 2H), 1.36 (s, 9H). HRMS (FAB+) calc'd for C<sub>46</sub>H<sub>47</sub>N<sub>11</sub>O<sub>10</sub>F<sub>5</sub> [M+H]<sup>+</sup> 1008.343, found 1008.342.

#### 4.4.4 Oligomerization procedure

A glass vial (1 dram) was charged with **4** (5.0 mg, 4.96  $\mu$ mole) and treated with a solution of CF<sub>3</sub>CO<sub>2</sub>H in CH<sub>2</sub>Cl<sub>2</sub> (1:1 CF<sub>3</sub>CO<sub>2</sub>H:CH<sub>2</sub>Cl<sub>2</sub>, 1 mL) and stirred at 23 °C for 10 min. The solvent was removed *in vacuo* and the residual solid was dried under high vacuum for 20 min. The solid was diluted with DMSO (500  $\mu$ L) followed by DIEA (80  $\mu$ L) and the solution was stirred at 23 °C for 20 h. After 20 h the reaction was complete by analytical HPLC analysis. The reaction was diluted to a final volume of 10 mL by addition of a solution of DMF in aqueous CF<sub>3</sub>CO<sub>2</sub>H (2:3 DMF:0.1% aqueous CF<sub>3</sub>CO<sub>2</sub>H). NOTE: A small amount of yellow insoluble material was observed and discarded. Purification by RP-HPLC yielded **1z** (13.9% yield), **2z** (5.5% yield), and **3z** (2.1% yield). The yield of **1z** is calculated from the mass of the purified and isolated material (0.5 mg). Yields for **2z** and **3z** were calculated based on **1z** using the relative product distribution as measured by integration of the preparative HPLC chromatogram at 310 nm (product distribution: 6.6:2.6:1.0 ratio of **1z**:**2z**:**3z**; UV integral values were normalized to the number of ImPyPyPy strands contained

in each cyclic oligomer). The benzyl carbamate (Cbz) protecting groups of **1z-3z** were removed as previously described.<sup>5g</sup> Characterization data for dimer **1** has been reported previously in Chapter 3 of this thesis.<sup>11</sup> Trimer **2** HRMS (MALDI-TOF) calc'd for C<sub>81</sub>H<sub>94</sub>N<sub>33</sub>O<sub>15</sub> [M+H]<sup>+</sup> 1768.7607, found 1768.7566. Tetramer **3** HRMS (MALDI-TOF) calc'd for C<sub>108</sub>H<sub>125</sub>N<sub>44</sub>O<sub>20</sub> [M+H]<sup>+</sup> 2358.0112, found 2358.0143.

#### 4.4.5 UV Absorption Spectrophotometry

Melting temperature analysis was performed on a Varian Cary 100 spectrophotometer equipped with a thermo-controlled cell holder possessing a cell path length of 1 cm. A degassed aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0 was used as analysis buffer. DNA duplexes and hairpin polyamides were mixed in 1:1 stoichiometry to a final concentration of 2 μM for each experiment. Prior to analysis, samples were heated to 90 °C and cooled to a starting temperature of 23 °C with a heating rate of 5 °C/min for each ramp. Denaturation profiles were recorded at λ = 260 nm from 23 °C to 90 °C with a heating rate of 0.5 °C/min. The reported melting temperatures were defined as the maximum of the first derivative of the denaturation profile.

#### 4.5 References and Notes

1. Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215–2235.
2. Dervan, P. B., and Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–299.
3. (a) Trauger, J. W., Baird, E. E., and Dervan, P. B. *Nature* **1996**, *382*, 559–561. (b) White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. *Nature* **1998**, *391*, 468–470. (c) Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104–109. (d) Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. *Science* **1998**, *282*, 111–115.
4. Hsu, C. F., Phillips, J. W., Trauger, J. W., Farkas, M. E., Belitsky, J. M., Heckel, A., Olenyuk, B. Z., Puckett, J. W., Wang, C. C. C., and Dervan, P. B. *Tetrahedron* **2007**, *63*, 6146–6151.
5. (a) Belitsky, J. M., Leslie, S. J., Arora, P. S., Beerman, T. A., and Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 3313–3318. (b) Crowley, K. S., Phillion, D. P., Woodard, S. S., Scheitzer, B. A., Singh, M., Shabany, H., Burnette, B., Hippenmeyer, P., Heitmeier, M., and Bashkin, J. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1565–1570. (c) Best, T. P., Edelson, B. S., Nickols, N. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12063–12068. (d) Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., and Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802–2818. (e) Xiao, X., Yu, P., Lim, H. S., Sikder, D., and Kodadek, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865–2868. (f) Nickols, N. G., Jacobs, C. S., Farkas, M. E., and Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 363–370. (g) Dose, C., Farkas, M. E., Chenoweth, D. M., and Dervan, P. B. *J. Am. Chem. Soc.* **2008**, *130*, 6859–6866. (h) Hsu, C. F., and Dervan, P. B. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5851–5855.

6. (a) Olenyuk, B. Z., Zhang, G. J., Klco, J. M., Nickols, N. G., Kaelin, Jr., W. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16768–16773. (b) Kageyama, Y., Sugiyama, H., Ayame, H., Iwai, A., Fujii, Y., Huang, L. E., Kizaka-Kondoh, S., Hiraoka, M., and Kihara, K. *Acta. Oncol.* **2006**, *45*, 317–324. (c) Nickols, N. G., Jacobs, C. S., Farkas, M. E., and Dervan, P. B. *ACS Chem. Biol.* **2007**, *2*, 561–571. (d) Nickols, N. G., and Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10418–10423. (e) Matsuda, H., Fukuda, N., Ueno, T., Tahira, Y., Ayame, H., Zhang, W., Bando, T., Sugiyama, H., Saito, S., Matsumoto, K., and others, O. *J. Am. Soc. Nephrol.* **2006**, *17*, 422–432. (f) Yao, E. H., Fukuda, N., Ueno, T., Matsuda, H., Matsumoto, K., Nagase, H., Matsumoto, Y., Takasaka, A., Serie, K., Sugiyama, H., and Sawamura, T. *Hypertension* **2008**, *52*, 86–92.
7. (a) Rucker V. C.; Foister S.; Melander C.; Dervan P. B. *J. Am. Chem. Soc.* **2003**, *125*, 1195–1202; (b) Chenoweth, D. M.; Viger, A.; Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2216–2217.
8. (a) Arndt, H. D.; Hauschild, K. E.; Sullivan, D. P.; Lake, K.; Dervan, P. B.; Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322–13323; (b) Kwonj, Y.; Arndt, H. D.; Qian, M.; Choi, Y.; Kawazoe, Y.; Dervan, P. B.; Uesugi, M. *J. Am. Chem. Soc.* **2004**, *126*, 15940–15941; (c) Hauschild, K. E.; Metzler, R. E.; Arndt, H. D.; Moretti, R.; Raffaele, M.; Dervan, P. B.; Ansari, A. Z. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5008–5013; (d) Stafford, R. L.; Arndt, H. D.; Brezinski, M. L.; Ansari, A. Z.; Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2660–2668; (e) Stafford, R. L.; Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 14026–14033; (f) Xiao, X.; Yu, P.; Lim, H. S.; Sikder, D.; Kodadek, T. *Angew. Chem. Int.* **2007**, *46*, 2865–2868.
9. (a) Cohen, J. D.; Sadowski, J. P.; Dervan, P. B. *Angew. Chem. Int. Ed.* **2007**, *46*, 7956–7959; (b) Schmidt, T. L.; Nandi, C. K.; Rasched, G.; Parui, P. P.; Brutschy, B.; Famulok, M.; Heckel, A. *Angew. Chem. Int. Ed.* **2007**, *46*, 4382–4384; (c) Cohen J. D.; Sadowski J. P.; Dervan P. B. *J. Am. Chem. Soc.* **2008**, *130*, 402–403.
10. Chenoweth, D. M.; Harki, D. A.; Dervan, P. B. *J. Am. Chem. Soc.* **2009**, In Press. (See Chapter 2 of this thesis)
11. Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B. *J. Am. Chem. Soc.* **2009**, In Press. (See Chapter 3 of this thesis)
12. Compound **1** targets the DNA sequence 5'-WGWWCW-3', where W =A/T or T/A.



## **Chapter 5: Allosteric Modulation of DNA by Small Molecules**

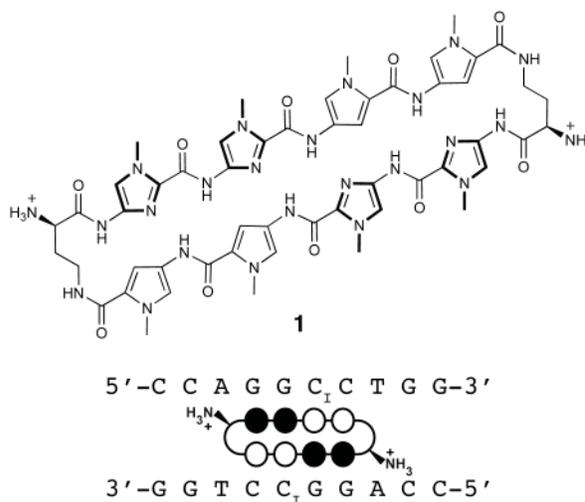
**Abstract**

Many human diseases are caused by dysregulated gene expression. The oversupply of one or more transcription factors may be required for the growth and metastatic behavior of human cancers. Cell permeable small molecules which can be programmed to disrupt transcription factor-DNA interfaces could silence aberrant gene expression pathways. Pyrrole-imidazole polyamides are DNA minor-groove binding molecules that are programmable for a large repertoire of DNA motifs. A high resolution X-ray crystal structure of an 8-ring cyclic Py/Im polyamide bound to the central six base pairs of the sequence d(5'-CCAGGC<sub>1</sub>CTGG-3')<sub>2</sub> reveals a 4 Å widening of the minor groove and compression of the major groove along with a >18° bend in the helix axis toward the major groove. This allosteric perturbation of the DNA helix provides a molecular basis for disruption of transcription factor-DNA interfaces by small molecules, a minimum step in chemical control of gene networks.

## 5.1 Introduction

Py/Im polyamides bind the minor groove of DNA sequence specifically,<sup>1,2</sup> encoded by side-by-side arrangements of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamide monomers. Im/Py pairs distinguish G•C from C•G base pairs, whereas Py/Py pairs are degenerate for T•A and A•T.<sup>3-6</sup> Antiparallel Py/Im strands are connected by  $\gamma$ -aminobutyric acid linker (GABA) to create a hairpin-shaped oligomer. Hairpin Py/Im polyamides have been programmed to bind a broad library of different DNA sequences.<sup>7</sup> They have been shown to permeate cell membranes,<sup>8-10</sup> access chromatin,<sup>11,12</sup> and disrupt protein-DNA interactions.<sup>2</sup> Disruption of transcription factor-DNA interfaces six bp in size such as HIF-1 $\alpha$ ,<sup>13-15</sup> androgen receptor (AR),<sup>16</sup> and AP-1<sup>17,18</sup> have been exploited for controlling expression of medically relevant genes such as VEGF, PSA, TGF- $\beta$ 1 and LOX-1 in cell culture experiments.<sup>13-18</sup> X-ray crystallography of antiparallel 2:1 binding polyamides in complex with DNA reveal a 1 Å widening of the minor groove.<sup>5,6</sup> This modest structural perturbation to the DNA helix by the side-by-side stacked arrangement of aromatic rings does not explain the large number of transcription factor-DNA interfaces disrupted by minor-groove binding hairpin Py/Im polyamides.<sup>2,5,6,19</sup> It must be that the turn unit in the hairpin oligomer connecting the two antiparallel strands plays a structural role.

Here we report the atomic resolution structure (1.18 Å resolution) of an 8-ring cyclic polyamide in complex with double helical DNA. The cyclic polyamide **1** is comprised of two antiparallel ImImPyPy strands capped by (*R*)- $\alpha$ -amino- $\gamma$  turn units. Polyamide **1**, which codes for the sequence 5'-WGGCCW-3' was co-crystallized with the palindromic DNA oligonucleotide sequence 5'-CCAGGC<sub>1</sub>CTGG-3' 10 base pairs in length (Figure 5.1). We observe significant structural allosteric perturbations of the DNA helix induced upon binding of GABA ( $\gamma$ -aminobutyric acid) turn-linked polyamides in the minor groove. A detailed view of the  $\alpha$ -amino- $\gamma$ -turn conformation and hydration

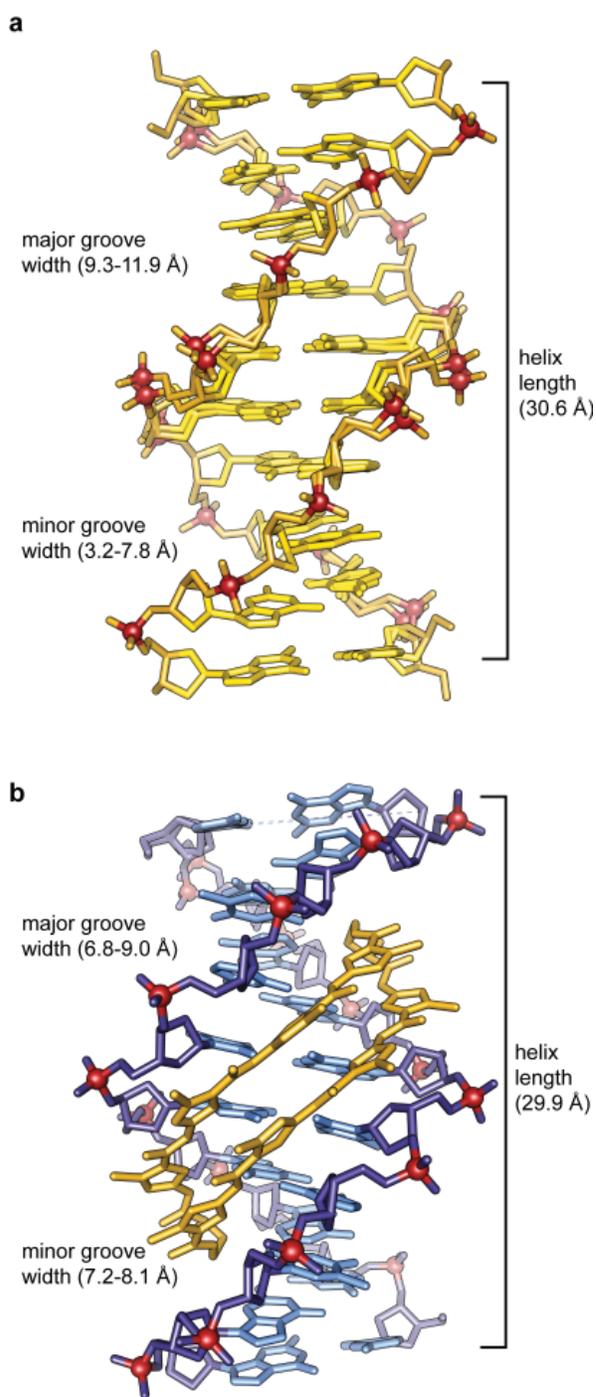


**Figure 5.1** Chemical structure of the cyclic polyamide and DNA sequence. Cyclic polyamide **1** targeting the sequence 5'-WGGCCW-3' shown with ball-and-stick model superimposed onto the DNA oligonucleotide used for crystallization. (Black circles represent imidazoles, open circles represent pyrroles, and ammonium substituted half circles at each end represent the (*R*)- $\alpha$ -amine- $\gamma$ -turn.

reveal a network of well-ordered water-mediated interactions between the polyamide and the minor groove floor of DNA.

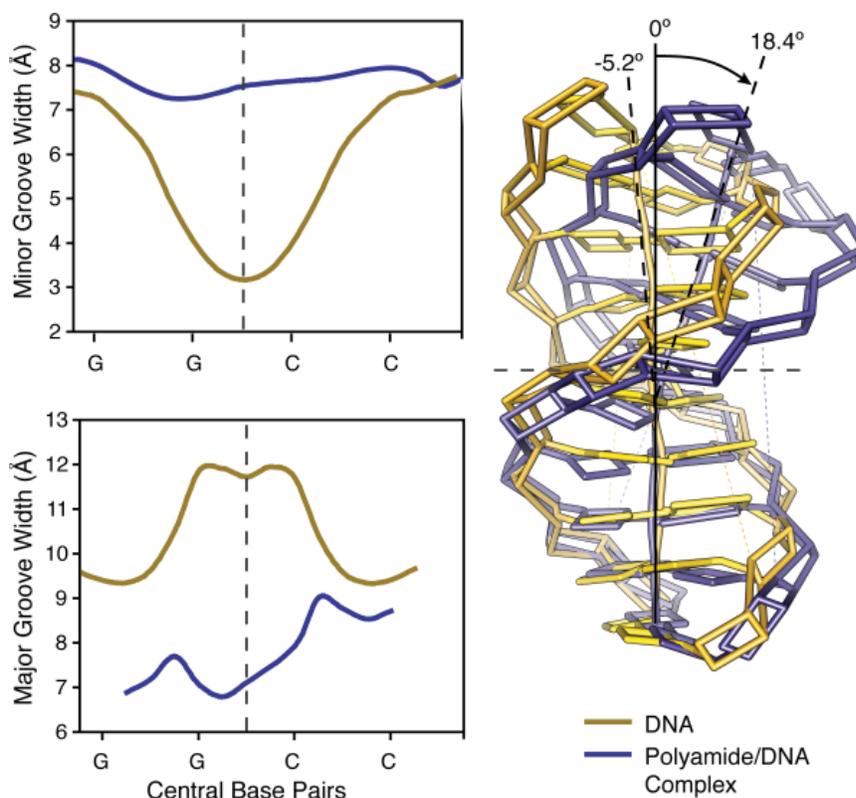
## 5.2 Results and Discussion

The structure was solved by direct methods to 1.18 Å resolution with synchrotron radiation. One cyclic polyamide bound to a single DNA duplex is present in the asymmetric unit of the crystal in the P1 space group. In the DNA complex the aromatic amino acids are bound with an N- to C- orientation of each ImImPyPy strand of the cycle adjacent to the 5' to 3' direction of the DNA. To assess DNA structural perturbations imposed by polyamide binding we compared our polyamide-DNA complex to the free DNA. We solved the X-ray structure of  $d(\text{CCAGGC}_1\text{CTGG})_2$  to a resolution of 0.98 Å for comparison.<sup>20</sup> Interestingly, this structure shows several discrete alternate conformations in 7 of the 10 nucleotides in each strand of the DNA duplex, illustrating the dynamic and conformationally mobile nature of the B-DNA sugar-phosphate backbone. The comparison of polyamide-DNA complex to free DNA is shown in Figure 5.2. Polyamide binding locks out the alternate DNA conformations, rigidifying the sugar-phosphate backbone, and strongly perturbing the overall helix structure. Binding of the polyamide widens the minor groove up to 4 Å while simultaneously compressing the major



**Figure 5.2** Comparison of native DNA to polyamide/DNA complex. a) Native DNA crystal structure at 0.98 Å resolution. b) Comparison to DNA/polyamide co-crystal structure at 1.18 Å resolution. (Both structure solved by direct methods.) showing the bound cyclic polyamide with electron density contoured at the 1.0  $\sigma$  level.

groove by 4 Å. The polyamide bends the DNA strand  $>18^\circ$  toward the major groove as shown in Figure 5.3, and shortening of the overall length of the helix by  $\sim 1$  Å (Figure 5.2).

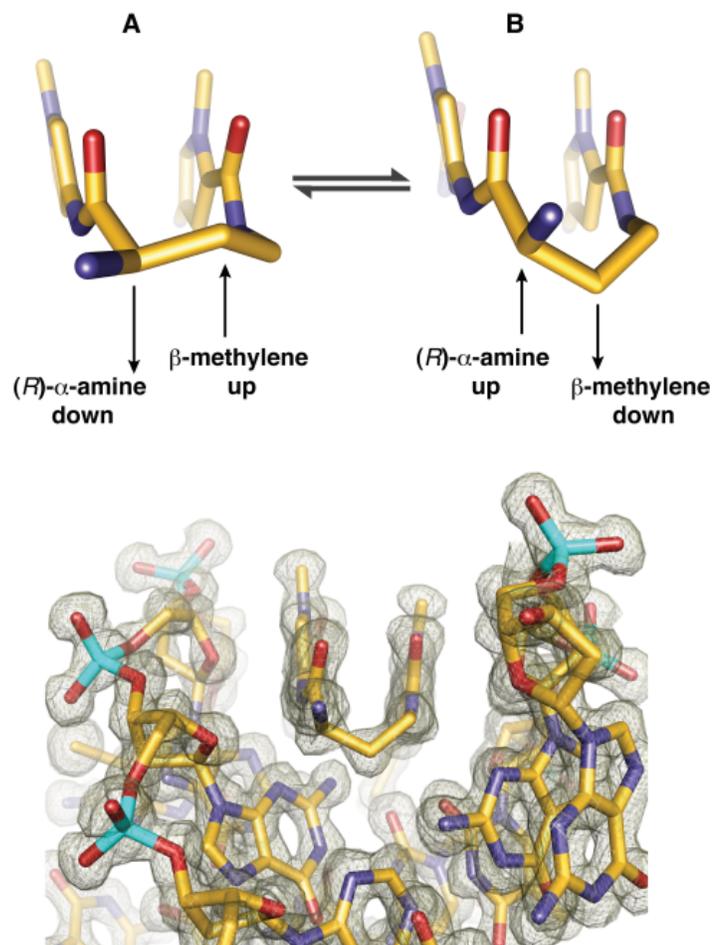


**Figure 5.3** Analysis of native DNA (yellow) compared to polyamide complexed DNA (blue). Chart on the top left shows variation in the minor groove width for native DNA (yellow) and polyamide-complexed DNA (blue) over the central core sequence 5'-GGCC-3'. Chart on the bottom left shows variation in the major groove width for native DNA (yellow) and polyamide complexed DNA (blue) over the central core sequence 5'-GGCC-3'. Overlay of the Curves calculated geometric helix model from each structure showing a DNA bend of  $> 18^\circ$  in the polyamide/DNA complex compared to native DNA.

Py/Im Polyamides linked by a GABA or substituted GABA can adopt either of two possible conformations shown in Figure 5.4. In conformation **A**, the amino group is directed toward the minor-groove wall of DNA with the potential for steric clash with the deoxyribose sugar. In conformation **B** the amine is directed up and out of the minor groove forcing the  $\beta$ -methylene to the floor of the minor groove with the potential for steric interaction with the nucleobases within van der Waals contact distance of the C2 hydrogen of adenine. We observe conformation **B** in our high resolution X-ray structure at both ends of the complex (Figure 5.4). It is possible that there is an intrinsic preference for conformation **A**, which relieves the  $\beta$ -methylene interaction with the floor of the minor groove.

However, for turn substitution at the  $\alpha$ -position interaction with the minor-groove wall becomes the dominant steric interaction, leading to conformational inversion. Figure 5.4 presents a view of the complex looking down the minor groove directly at the polyamide turn linkage. From this view, significant van der Waals interactions can be observed between the outside face of the pyrrole-imidazole strands and the walls of the minor groove, which form a deep binding pocket for the cycle. Approximately 40% of the polyamide surface area is buried leaving only the top of the methyl groups on the heterocycles, the amide carbonyl oxygens, and the chiral  $\alpha$ -ammonium turn solvent exposed. In addition the turn unit introduces conformational constraints preventing slipped or linear binding modes.

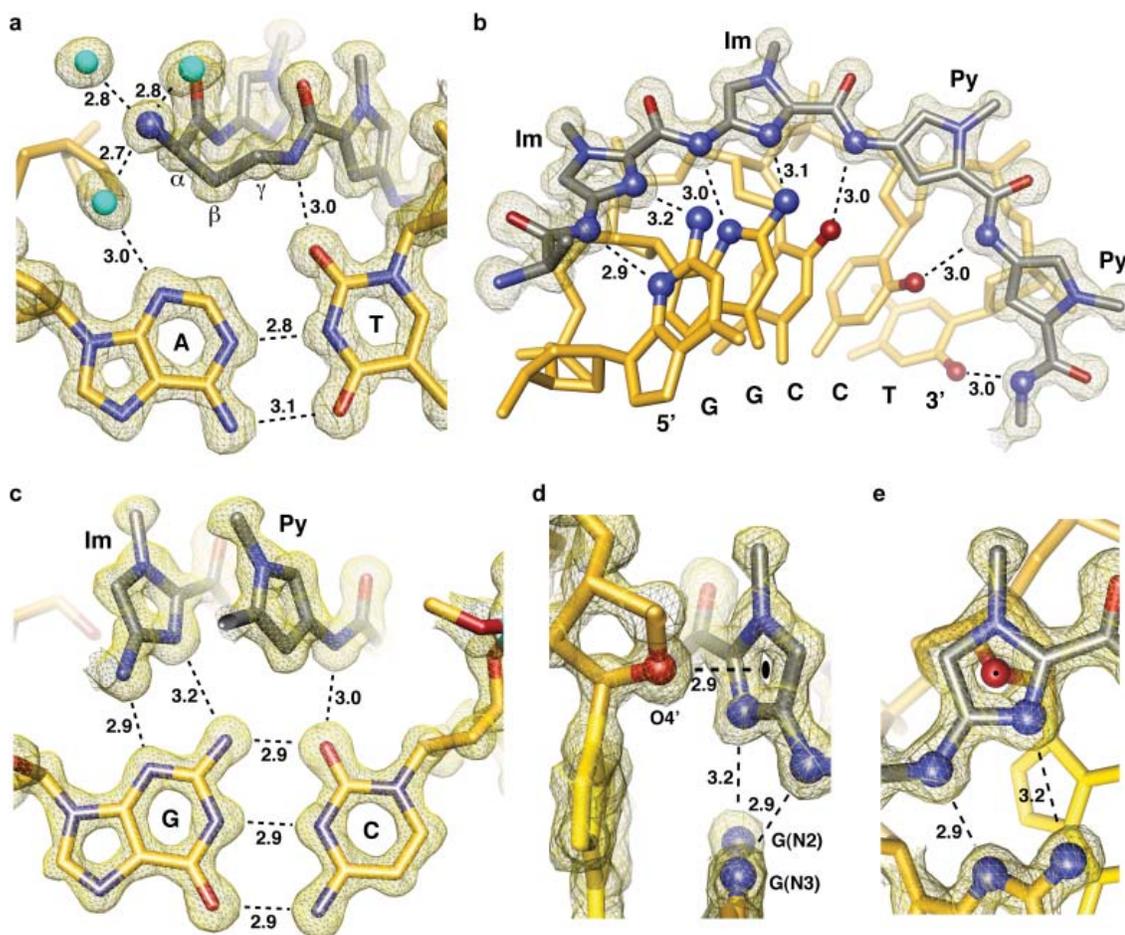
The conformational constraints imposed by the turn linkage result in ring placement that is an intermediate of ring-over-ring and ring-over-amide between adjacent PyPyImIm strands. This alignment allows the ring pairs to remain in phase with the nucleobases as the polyamide adopts an isohelical conformation complementary to the DNA helix. This is highlighted by comparison to the 2:1 structure in which the rings lie over the carboxamide linkages of the adjacent strand.<sup>5</sup> The conformational constraints



**Figure 5.4** Conformation of the  $\alpha$ -amino substituted GABA turn. (top), Two possible Conformations **A** and **B** are shown with conformation **A** directing the  $\beta$ -methylene up and away from the minor groove floor while orienting the  $\alpha$ -ammonium toward the minor groove wall. Conformation **B** presents the  $\beta$ -methylene down toward the minor-groove floor while orienting the  $\alpha$ -ammonium up and out of the minor-groove, relieving possible steric interaction with the sugar-phosphate backbone (minor-groove wall). (bottom), View looking down the DNA minor-groove, showing the  $(R)$ - $\alpha$ -amine- $\gamma$ -turn conformation observed in the X-ray crystal structure, which matches that of conformation **B**. Electron density map is contoured at the  $1.0 \sigma$  level.

imposed by the turn and inability of the ligand to slip into a possibly more preferred orientation may impact the overall DNA structure by inducing bending and other allosteric distortions accommodated by the plasticity of DNA. In addition, the turn-constrained cycle may have a major entropic driving force leading to substantial pre-organization, increased affinity, and increased specificity by locking out unproductive conformations and alternate binding modes. The van der Waals interactions between rings may also lead to cooperativity in the binding process. In addition, we find a shell of highly ordered water molecules around the  $\alpha$ -ammonium substituent and a water-mediated hydrogen bond from the ammonium to the N3 lone-pair of the adenine under the turn. The hydration pattern around the turn is highly conserved at both ends of the structure and the water mediated hydrogen bonds are within  $\sim 2.7$ – $2.9$  Å from the ammonium to water to the adenine lone-pair (Figure 5.5a).

The amide NH's and imidazole lone-pairs form a continuous series of direct hydrogen bonds to the floor of the DNA minor-groove, while the imidazoles impart specificity for the exocyclic amine of guanine through relief of a steric interaction and a G(N2-hydrogen)-Im (lone pair) hydrogen bond. The amides linking the aromatic rings and the turns contribute hydrogen bonds to the purine N3 and pyrimidine O2 lone pairs. All amides are within hydrogen bonding distance of a single DNA base ( $\sim 3.0$  Å average, Figure 5.9). In all there are 10 direct amide hydrogen bonds (average distance =  $2.97$  Å), 4 direct imidazole hydrogen bonds (2 terminal average distance =  $3.27$  Å and 2 internal average distance =  $3.05$  Å), and 2 (*R*)- $\alpha$ -ammonium turn water-mediated hydrogen bonds (average distance amine to water =  $2.75$  Å and average distance from the water to adenine N3 =  $2.98$  Å) to the floor of the DNA minor groove with at least one interaction for all 12 DNA base-pairs in the 6 bp binding site for a total of 16 hydrogen bond interactions between the cyclic polyamide and the floor of the DNA minor-groove. These 16 hydrogen bonds utilize every heteroatom of the polyamide presented to the floor of the DNA minor-groove, which exactly matches the total number of Watson-Crick hydrogen bonds between all the DNA base pairs in the 6 bp binding site. In addition to these 16 hydrogen bonds, we find unique weak interactions in the form of lone pair- $\pi$  interactions<sup>21,22</sup> between the center of the leading imidazole ring and the lone pair of the adjacent deoxyribose O4' oxygen (Figure 5.5d and 5.5e). Interestingly this interaction is only observed for the terminal imidazole aromatic ring and analysis of qualitative electrostatic potential surfaces substantiates the electropositive nature of the imidazole under these conditions (Figure 5.13).<sup>23</sup>



**Figure 5.5** Direct and water-mediated non-covalent molecular recognition interactions. a) Geometry of the  $\alpha$ -amino turn interacting with the AT base pair through water-mediated hydrogen bonding interactions. Structural basis for the turn preference for AT versus GC is demonstrated by the  $\beta$ -methylene conformational preference, which points down toward the DNA minor-groove floor within van der Waals contact distance of the adenine base. b) Isolated view of one half of the macrocyclic-polyamide showing hydrogen bond distances made to the DNA minor groove floor by the imidazoles and amides of compound **1**. c) Im-Py pair showing the mechanism for GC specificity. d) Interaction of the O4' oxygen of a deoxyribose sugar with the terminal imidazole aromatic ring through a lone pair- $\pi$  interaction. The sugar conformation is C2'-endo at the N-terminal imidazole of the polyamide with the sugar oxygen lone-pair pointing directly to the centroid of the imidazole ring. The distance between the sugar oxygen and the ring centroid is 2.90 Å, which is less than the sum of the van der Waals radii to any atom in the imidazole ring. Electrostatic potential maps calculated at the HF/3-21g\* level of theory show the slightly electropositive nature of the imidazole ring under these conditions (Figure 5.13). e) View of the O4' deoxyribose oxygen atom looking through the imidazole ring showing the ring centroid superimposed on the oxygen atom. All distances are reported in angstroms (Å) and all electron density maps are contoured at the 1.0  $\sigma$  level (Im = imidazole and Py = pyrrole).

### 5.3 Conclusion

The crystal structure presented highlights the molecular recognition of turn-linked polyamides in the minor-groove of DNA and provides insight into the allosteric modulation of

B-form DNA by hairpin oligomers. The DNA structural distortion induced upon polyamide minor-groove binding provides an allosteric model for disrupting DNA:transcription factor interfaces in the promoters of selected genes. The ability of DNA to undergo short and long-range allosteric effects coupled with DNA binding by proteins can have influence over important processes such as modulation of eukaryotic gene networks.<sup>24-27</sup> Allosteric communication along and through the DNA helix forms the basis for cooperative interactions among transcription factor regulatory networks such as the interferon- $\beta$  enhanceosome.<sup>25</sup> The potential for allosteric control over the transcriptional machinery provides a powerful concept for the design of small molecules that can bind to distinct locations on DNA with the possibility of modulating transcription factor activity.<sup>1,2,19,27</sup>

## 5.4 Experimental Section

### 5.4.1 General

Chemicals and solvents were purchased from Sigma-Aldrich and Hampton Research and were used without further purification. Water (18 M $\Omega$ ) was purified using a Millipore MilliQ purification system. Analytical HPLC analysis was conducted on a Beckman Gold instrument equipped with a Phenomenex Gemini analytical column (250 x 4.6 mm, 5  $\mu$ M), a diode array detector, and the mobile phase consisted of a gradient of acetonitrile (MeCN) in 0.1% (v/v) aqueous trifluoroacetic acid (TFA). Preparative HPLC was performed on an Agilent 1200 system equipped with a solvent degasser, diode array detector, and a Phenomenex Gemini column (5  $\mu$ m particle size, C18 110A, 250 x 21.2 mm, 5 micron). A gradient of MeCN in 0.1% (v/v) aqueous trifluoroacetic acid (TFA) was utilized as the mobile phase. UV-Vis measurements were made on a Hewlett-Packard diode array spectrophotometer (model 8452 A) and polyamide concentrations were measured in 0.1% (v/v) aqueous TFA using an extinction coefficient of 69200 M<sup>-1</sup>cm<sup>-1</sup> at  $\lambda_{\text{max}}$  near 310 nm. Matrix-assisted LASER desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on an Applied Biosystems Voyager DR Pro spectrometer using  $\alpha$ -cyano-4-hydroxycinnamic acid as matrix.

### 5.4.2 Synthesis

Polyamide **1** was synthesized by standard solid-phase synthesis methods<sup>28,29</sup> on oxime resin (Figure 5.6) and purified by reverse-phase high-performance liquid chromatography (Figure 5.7).

### 5.4.3 Cyclo-(-ImImPyPy-(R) <sup>$\alpha$</sup> -BocHN $\gamma$ -ImImPyPy-(R) <sup>$\alpha$</sup> -H<sub>2</sub>N $\gamma$ -) (6)

Oxime resin (**R3**) was generated by manual solid-phase synthesis from Kaiser oxime resin (1 g, 0.48 mmol/g, Novabiochem) using previously described Boc-protected monomers.<sup>28,29</sup> Boc-Py-OBt (**2**) (716 mg, 2 mmol) was dissolved in 2 mL of DMF and added to 1 g of oxime resin followed by 1 mL of DIEA. The reaction was left in a 37 °C shaker for 12 h. The resin was drained, washed with DMF 3x, DCM 3x, and the Boc group was removed upon treatment with 20% TFA/DCM for 30 min. After draining the resin and washing with DCM 3x followed by DMF 3x the second pyrrole residue was coupled in the same fashion as the first, with complete coupling after 2 h at 23 °C. Boc-Im-OH (**3**) (482 mg, 2 mmol) was dissolved in 2 mL of DMF and treated with 1.14 g (2 mmol) of PyBOP and 2 mL of DIEA. This solution was stirred for 5 minutes prior to addition to the resin vessel. Coupling was allowed to proceed for 2 h at 23 °C. The Boc-Im residue was deprotected using a 50% TFA/DCM solution for 30 min at room temperature followed by draining the resin and washing with DCM 3x and DMF 3x. A second imidazole residue was coupled following the exact same procedure as the first. The turn unit,  $\alpha$ -Fmoc- $\gamma$ -Boc-(*R*)-diaminobutyric acid (Fmoc-D-Dab-(Boc)-OH) (**4**) (660 mg, 1.5 mmol) was activated with PyBOP (855 mg, 1.5 mmol) in 2 mL of DMF and 1 mL of DIEA at 23 °C for 15 min prior to addition to the resin. Coupling was allowed to proceed for 2 h at 37 °C. After deprotection with 20% TFA/DCM for 30 min, the next two pyrrole residues were attached in exactly the same manner as previously described using Boc-Py-OBt. The last two imidazoles were added in the same fashion as the previous two. The final turn unit,  $\alpha$ -Boc- $\gamma$ -Fmoc-(*R*)-diaminobutyric acid (Boc-D-Dab-(Fmoc)-OH) (**5**) (660 mg, 1.5 mmol) was activated with PyBOP (855 mg, 1.5 mmol) in 2 mL of DMF and 1 mL of DIEA at 23 °C for 15 min prior to addition to the resin. Coupling was allowed to proceed for 2 h at 37 °C. After Fmoc deprotection with 25% piperidine/DMF for 3x5 min, the resin was washed with DMF 6x and 1:1 DMF/DIEA 3x. Next, the resin was diluted with 10 mL of DMF and stored in a 37 °C shaker on medium speed for 24 h. The resin was filtered off and the DMF concentrated to 1 mL volume, taken up in 9 mL of H<sub>2</sub>O (0.1% TFA), and purified by preparative reverse-phase HPLC to give *cyclo*-(-ImImPyPy-(*R*) <sup>$\alpha$ -BocHN $\gamma$</sup> -ImImPyPy-(*R*) <sup>$\alpha$ -H<sub>2</sub>N $\gamma$</sup> -) (**6**) as a fluffy white solid in 0.1% overall yield (0.480  $\mu$ mol). *Cyclo*-(-ImImPyPy-(*R*) <sup>$\alpha$ -BocHN $\gamma$</sup> -ImImPyPy-(*R*) <sup>$\alpha$ -H<sub>2</sub>N $\gamma$</sup> -) (**6**) MALDI-TOF MS (*m/z*): calc'd for C<sub>57</sub>H<sub>68</sub>N<sub>24</sub>NaO<sub>12</sub> [M+Na]<sup>+</sup> 1303.53, found 1303.36.

#### 5.4.4 *Cyclo*-(-ImImPyPy-(*R*) <sup>$\alpha$ -H<sub>2</sub>N $\gamma$</sup> -ImImPyPy-(*R*) <sup>$\alpha$ -H<sub>2</sub>N $\gamma$</sup> -) (**1**)

A solution of *cyclo*-(-ImImPyPy-(*R*) <sup>$\alpha$ -BocHN $\gamma$</sup> -ImImPyPy-(*R*) <sup>$\alpha$ -H<sub>2</sub>N $\gamma$</sup> -) (**6**) (0.400  $\mu$ mol) in anhydrous TFA/DCM (1:1, 500  $\mu$ L) was stirred at 23 °C for 5 min prior to being taken up in 9.5 mL of H<sub>2</sub>O

(0.1% TFA), and purified by preparative reverse-phase HPLC to give *cyclo*-(-ImImPyPy-(R)<sup>α-H<sub>2</sub>N</sup>γ-ImImPyPy-(R)<sup>α-H<sub>2</sub>N</sup>γ-) **1** as a fluffy white solid in 90% yield (0.360 μmol). *Cyclo*-(-ImImPyPy-(R)<sup>α-H<sub>2</sub>N</sup>γ-ImImPyPy-(R)<sup>α-H<sub>2</sub>N</sup>γ-) **1** MALDI-TOF MS (*m/z*): calc'd for C<sub>52</sub>H<sub>61</sub>N<sub>24</sub>O<sub>10</sub> [M+H]<sup>+</sup> 1181.50, found 1181.55; calc'd for C<sub>52</sub>H<sub>60</sub>N<sub>24</sub>O<sub>10</sub>Na [M+Na]<sup>+</sup> 1203.48, found 1203.37; calc'd for C<sub>52</sub>H<sub>60</sub>N<sub>24</sub>O<sub>10</sub>K [M+K]<sup>+</sup> 1219.46, found 1219.33.

#### 5.4.5 Oligonucleotide Purification and Crystallization

Oligonucleotides were purchased HPLC purified from Trilink Biotechnologies (San Diego, CA). Prior to use, oligonucleotides were de-salted using a Waters Sep-Pak cartridge (5g, C-18 sorbent). The Sep-Pak was pre-washed with acetonitrile (25 mL, 3x) followed by MilliQ water (25 mL, 3x). The oligonucleotide was dissolved in 5 mL of 2.0 M NaCl and loaded directly onto the sorbent followed by a wash with 5 mL of 2.0 M NaCl and 250 mL of MilliQ water. The oligonucleotide was eluted with acetonitrile:water (1:1) and lyophilized to dryness. Single strand DNA was quantitated by UV-Vis spectroscopy and incubated with a 2:1 ratio of DNA to polyamide prior to crystallization. Crystals were obtained after 2-8 weeks from a solution of 0.5 mM duplex DNA, 0.65 mM polyamide, 21% 2-methyl-2,4-pentanediol (MPD), 35 mM calcium acetate, 10 mM Tris pH 7.5 equilibrated in sitting drops against a reservoir of 35% MPD at 4 °C. Crystals were collected in Hampton nylon CryoLoops (10 micron, 0.1 mm) and flash cooled to 100 K prior to data collection (Figure 5.8).

#### 5.4.6 UV-visible analysis

DNA/polyamide complex formation was verified prior to structure solution by UV-Visible spectroscopy. Crystals were collected in Hampton nylon CryoLoops (10 micron, 0.1 mm) and washed with crystallization buffer 3 times prior to dissolution in 50 μL of MilliQ water. UV-Visible spectroscopy of the dissolved crystals confirmed the presence of polyamide and DNA duplex in a 1:1 stoichiometry (Figure 5.8).

#### 5.4.7 Data collection, Structure Determination, and Refinement

Polyamide-DNA crystals grew in space group P1 with unit cell dimensions  $a = 22.50$ ,  $b = 25.14$ ,  $c = 29.09$ ,  $\alpha = 66.53$ ,  $\beta = 79.28$ ,  $\gamma = 79.57$ , and one polyamide-duplex DNA complex in the asymmetric unit. This data set was collected at Stanford Synchrotron Radiation Laboratory (SSRL) beamline 12-2 with a MAR Research imaging plate detector at wavelength 0.97 Å. DNA only crystals grew in space group C2 (C 1 2 1) with unit cell dimensions  $a = 31.827$ ,  $b = 25.636$ ,  $c = 34.173$ ,  $\alpha = 90.00$ ,

$\beta = 116.72$ ,  $\gamma = 90.00$  and one DNA strand in the asymmetric unit. This data set was collected at Stanford Synchrotron Radiation Laboratory (SSRL) beamline 11-1 with a MAR Research imaging plate detector at wavelength 0.999 Å (Table 5.1).

Data was processed with MOSFLM<sup>30</sup> and SCALA<sup>31</sup> from the CCP4 suite of programs.<sup>31</sup> Both crystals were solved by direct methods using the SHELX suite of programs (SHELXD).<sup>32,33</sup> Model building and structure refinement was done with Coot<sup>34</sup> and REFMAC5.<sup>35</sup> The final polyamide-DNA complex was refined to an R factor of 9.8 % and an  $R_{\text{free}}$  of 13.6 %. The final DNA structure was refined to an R factor of 10.9 % and an  $R_{\text{free}}$  of 14.3 %. Anisotropic B factors were refined in the final stages and riding hydrogens included (Table 5.1).

#### 5.4.8 Structure Analysis and Figure Preparation

DNA helical parameters were calculated using the program Curves and 3DNA.<sup>36,37</sup> Molecular electrostatic potential maps were calculated at the HF/3-21g\* level using the Gamess program (Figure 5.13).<sup>38-40</sup> Distance measurements and least squares fitting procedures for ring-centroid measurements were performed using UCSF Chimera<sup>41</sup> and Mercury.<sup>42</sup> Structural figures were prepared using UCSF Chimera.<sup>41</sup>

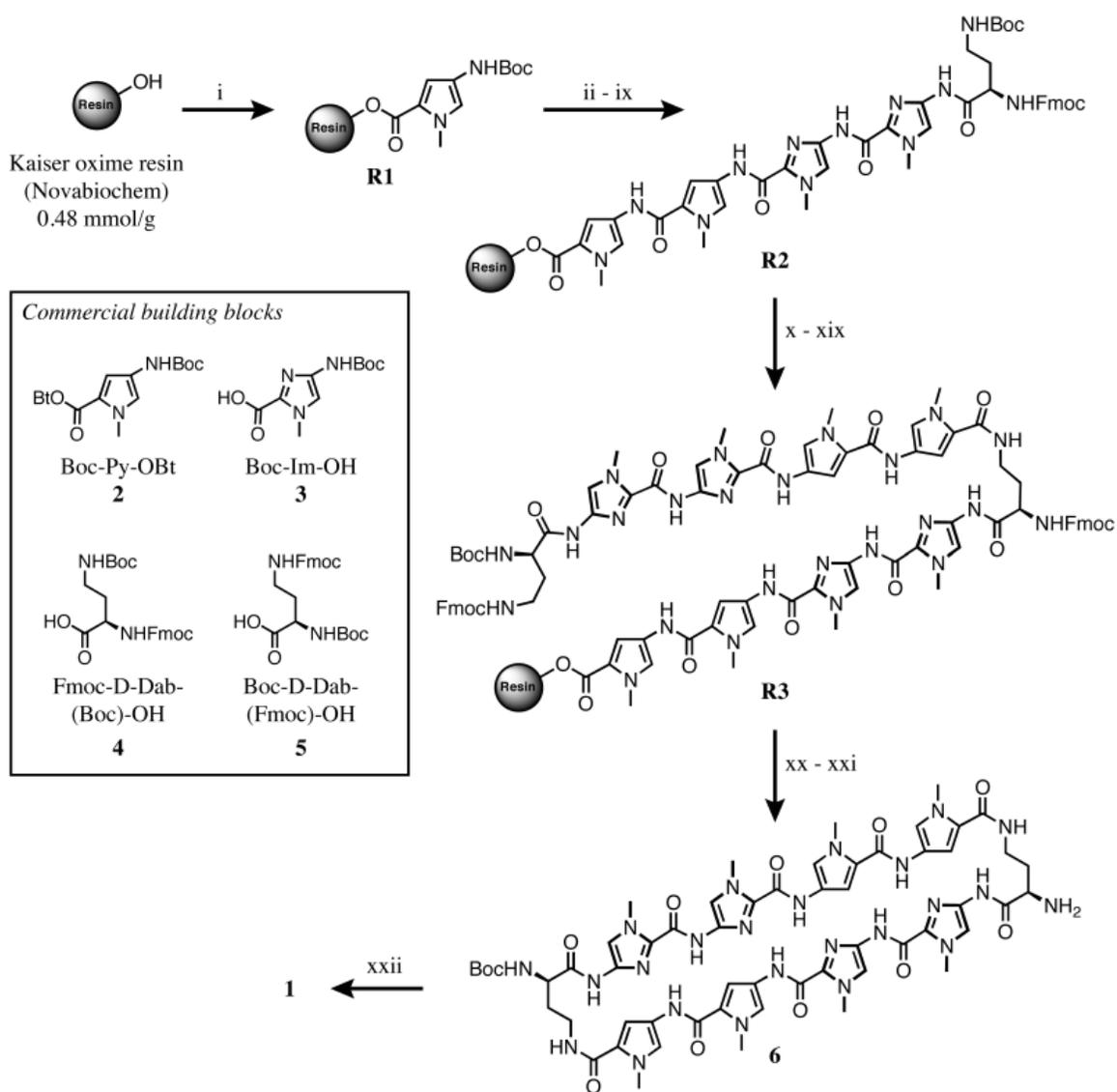
### 5.5 Notes and References

1. Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg Med Chem* **9**, 2215-2235 (2001).
2. Dervan, P. B. & Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Current Opinion in Structural Biology* **13**, 284-299 (2003).
3. Trauger, J. W., Baird, E. E. & Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559-561 (1996).
4. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-470 (1998).
5. Kielkopf, C. L., Baird, E. E., Dervan, P. B. & Rees, D. C. Structural basis for G•C recognition in the DNA minor groove. *Nat Struct Biol* **5**, 104-109 (1998).
6. Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., *et al.* A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science* **282**, 111-115 (1998).
7. Hsu, C. F., Phillips, J. W., Trauger, J. W., Farkas, M. E., *et al.* Completion of a Programmable DNA-Binding Small Molecule Library. *Tetrahedron* **63**, 6146-6151 (2007).
8. Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., *et al.* Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res* **32**, 2802-2818 (2004).

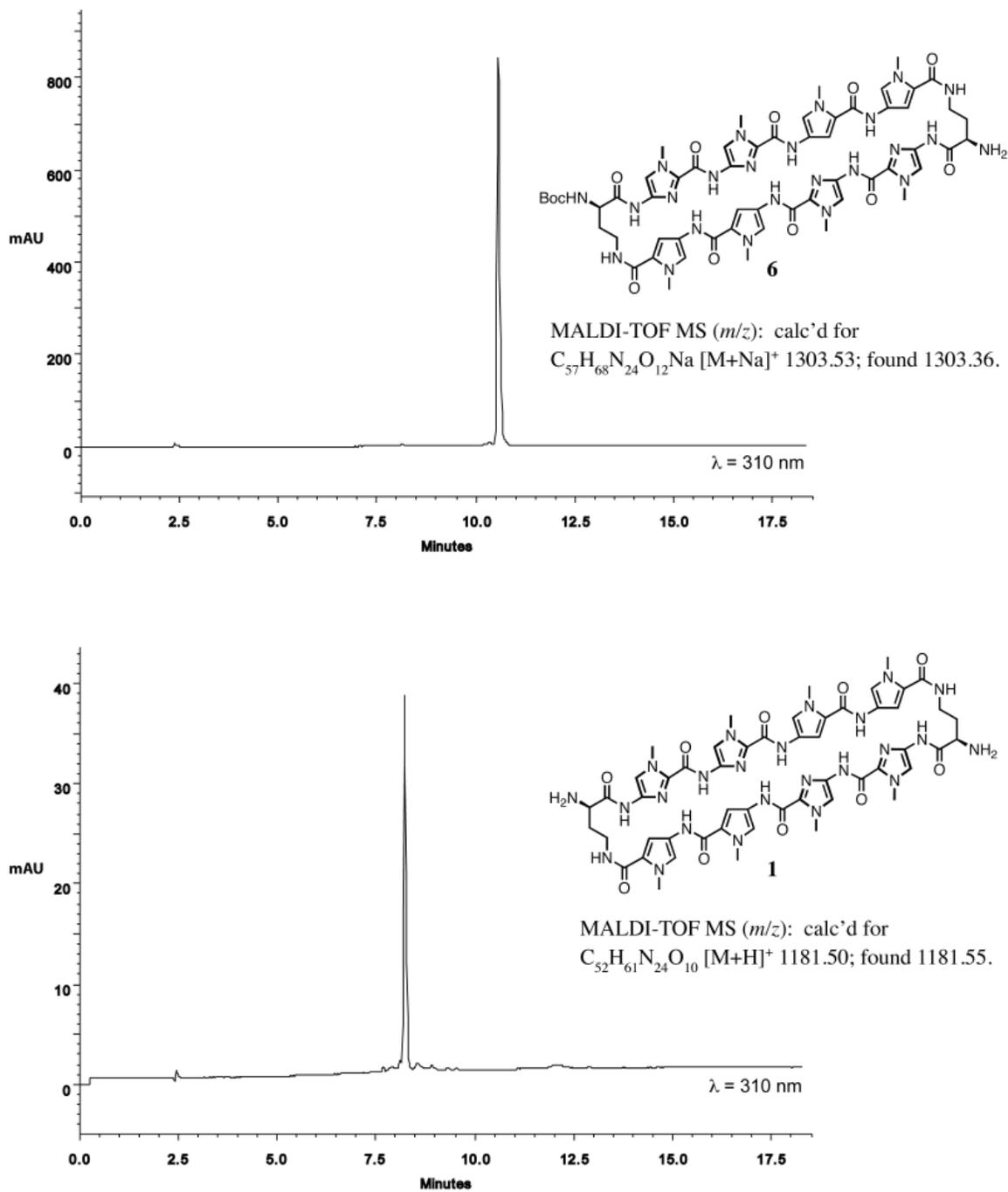
9. Nickols, N. G., Jacobs, C. S., Farkas, M. E. & Dervan, P. B. Improved nuclear localization of DNA-binding polyamides. *Nucleic Acids Res* **35**, 363-370 (2007).
10. Hsu, C. F. & Dervan, P. B. Quantitating the concentration of Py-Im polyamide-fluorescein conjugates in live cells. *Bioorg Med Chem Lett* **18**, 5851-5855 (2008).
11. Gottesfeld, J. M., Melander, C., Suto, R. K., Raviol, H., Luger, K., and Dervan, P. B. Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides. *J. Mol. Biol.* **309**, 615-629 (2001).
12. Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* **326**, 371-380 (2003).
13. Olenyuk, B. Z., Zhang, G. J., Klco, J. M., Nickols, N. G., *et al.* Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc Natl Acad Sci U S A* **101**, 16768-16773 (2004).
14. Kageyama, Y., Sugiyama, H., Ayame, H., Iwai, A., *et al.* Suppression of VEGF transcription in renal cell carcinoma cells by pyrrole-imidazole hairpin polyamides targeting the hypoxia responsive element. *Acta Oncol* **45**, 317-324 (2006).
15. Nickols, N. G., Jacobs, C. S., Farkas, M. E. & Dervan, P. B. Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface. *ACS Chem Biol* **2**, 561-571 (2007).
16. Nickols, N. G. & Dervan, P. B. Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc Natl Acad Sci USA* **104**, 10418-10423 (2007).
17. Matsuda, H., Fukuda, N., Ueno, T., Tahira, Y., *et al.* *Journal of the American Society of Nephrology* **17**, 422-432 (2006).
18. Yao, E. H., Fukuda, N., Ueno, T., Matsuda, H., *et al.* Novel gene silencer pyrrole-imidazole polyamide targeting lectin-like oxidized low-density lipoprotein receptor-1 attenuates restenosis of the artery after injury. *Hypertension* **52**, 86-92 (2008).
19. Nguyen-Hackley, D. H., Ramm, E., Taylor, C. M., Joung, J. K., *et al.* Allosteric inhibition of zinc-finger binding in the major groove of DNA by minor-groove binding ligands. *Biochemistry* **43**, 3880-3890 (2004).
20. Heinemann, U., and Alings, C. Crystallographic study of one turn of G/C-rich B-DNA. *J. Mol. Biol.* **210**, 369-381 (1989).
21. Egli, M., and Sarkhel, S. Lone Pair-Aromatic Interactions: To Stabilize or Not to Stabilize *Acc. Chem. Res.* **40**, 197-205 (2007).
22. Gallivan, J. P., and Dougherty, D. A. Can Lone Pairs Bind to a  $\pi$  System? The Water-Hexafluorobenzene Interaction *Org. Lett.* **1**, 103-106 (1999).
23. Mecozzi, S., West, A. P., and Dougherty, D. A. *Proc. Natl. Acad. Sci. USA.* **93**, 10566-10571 (1996).
24. Hogan, M., Dattagupta, N., and Crothers, D. M. Transmission of allosteric effects in DNA *Nature* **278**, 521-524 (1979).
25. Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111-1123 (2007).

26. Lavelle, C. DNA torsional stress propagates through chromatin fiber and participates in transcriptional regulation. *Nat. Struct. Mol. Biol.* **15**, 146-154 (2008).
27. Moretti, R., Donato, L. J., Brezinski, M. L., Stafford, R. L., *et al.* Targeted chemical wedges reveal the role of allosteric DNA modulation in protein-DNA assembly. *ACS Chem Biol* **3**, 220-229 (2008).
28. Baird, E. E., Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146 (1996).
29. Belitsky, J. M., Nguyen, D. H., Wurtz, N. R., Dervan, P. B. Solid-phase synthesis of DNA binding polyamides on oxime resin. *Bioorg. Med. Chem.* **10**, 2767-2774 (2002).
30. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**, (1992).
31. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763 (1994).
32. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112-122 (2008).
33. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772-1779 (2002).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255 (1997).
36. Lavery, R., Sklenar, H. Curves 5.2: Helical analysis of irregular nucleic acids. *Biochimie Theorique, CNRS URA*, **77** (1997).
37. Lu, X. J., Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids. Res.*, **31**, 5108-5121 (2003).
38. Schmidt, M. W., Baldrige, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., *et al.* General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, **14**, 1347-1363 (1993).
39. Binkley, J. S., Pople, J. A., & Hehre, W. J. Self-Consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *Journal of the American Chemical Society*, **102**, 939-947 (1980).
40. Pietro, W. J., Francl, M. M., Hehre, W. J., DeFrees, D. J., Pople, J. A., & Binkley, J. S. Self-Consistent molecular orbital methods. 24. Supplemented small split-valence basis sets for second-row elements. *Journal of the American Chemical Society*, **104** 5039-5048 (1982).
41. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., *et al.* UCSF chimera-a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**, 1605-12 (2004).
42. Macrae, C.F., Edgington, P.R., McCabe, P., Pidcock, E., Shields, G.P., Taylor, R., Towler, M., van de Streek, J. Mercury: visualization and analysis of crystal structures. *J. Appl. Cryst.* **39**, 453-457, (2006).

## 5.6 Spectra, Data Statistics, and Supplemental Information



**Figure 5.6** Solid-phase synthesis of cyclic polyamide **1** using Kaiser oxime resin and the commercially available building blocks presented above. Reagents and conditions: (i) Boc-Py-OBt **2**, DIEA, DMF; (ii) 20% TFA/DCM; (iii) Boc-Py-OBt **2**, DIEA, DMF; (iv) 20% TFA/DCM; (v) Boc-Im-OH **3**, PyBOP, DIEA, DMF; (vi) 50% TFA/DCM; (vii) Boc-Im-OH **3**, PyBOP, DIEA, DMF; (viii) 50% TFA/DCM; (ix) Fmoc-D-Dab(Boc)-OH **4**, PyBOP, DIEA, DMF; (x) 20% TFA/DCM; (xi) Boc-Py-OBt **2**, DIEA, DMF; (xii) 20% TFA/DCM; (xiii) Boc-Py-OBt **2**, DIEA, DMF; (xiv) 20% TFA/DCM; (xv) Boc-Im-OH **3**, PyBOP, DIEA, DMF; (xvi) 50% TFA/DCM; (xvii) Boc-Im-OH **3**, PyBOP, DIEA, DMF; (xviii) 50% TFA/DCM; (xix) Boc-D-Dab-(Fmoc)-OH **5**, PyBOP, DIEA, DMF; (xx) piperidine, DMF; (xxi) DMF, 37°C, 24 h; (xxii) 50% TFA/DCM.



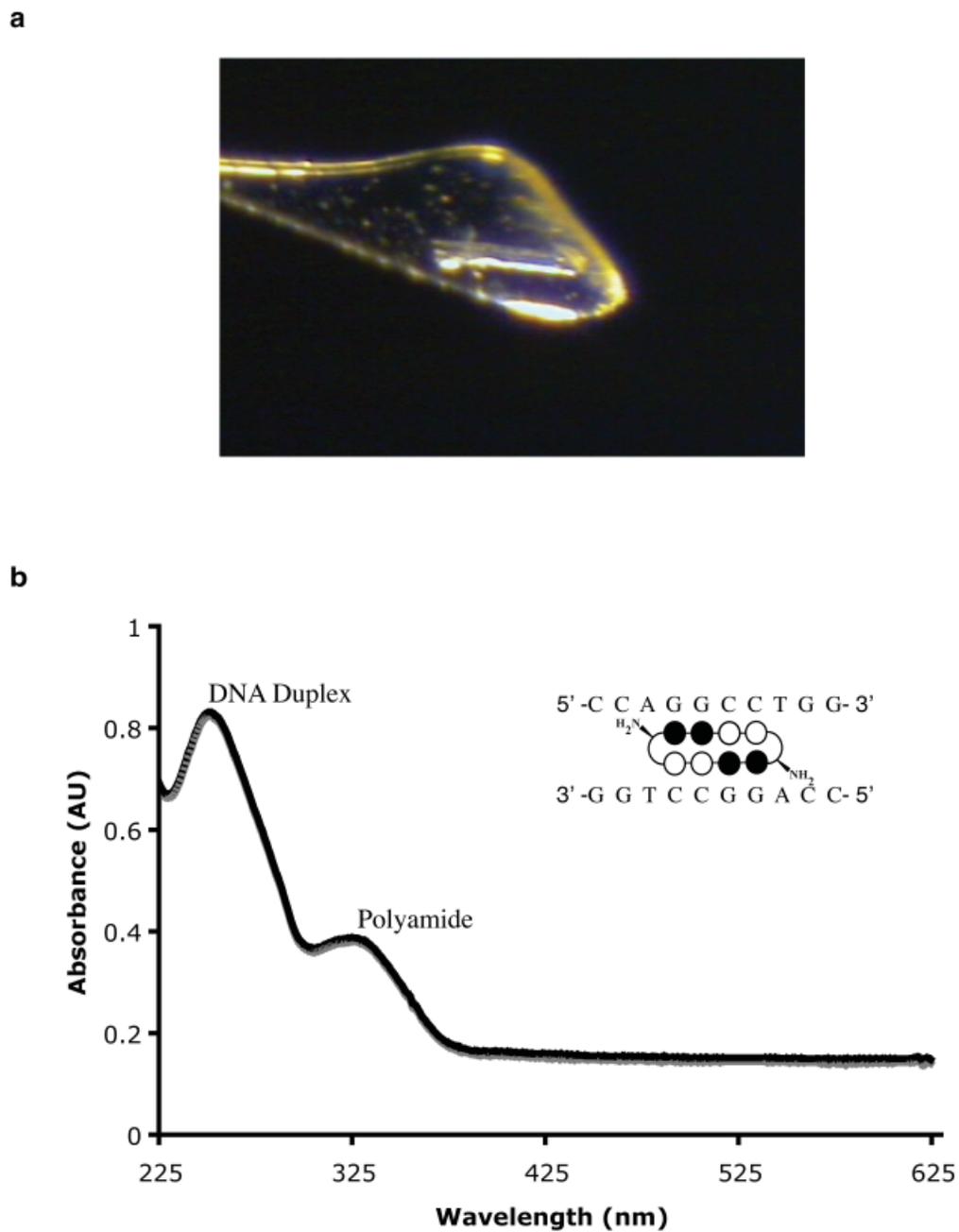
**Figure 5.7** Polyamide analytical data. Analysis of polyamides **6** and **1** by analytical RP-HPLC and MALDI-TOF MS. Wavelength shown is at 310 nm.

**Table 5.1** Data collection and refinement statistics.

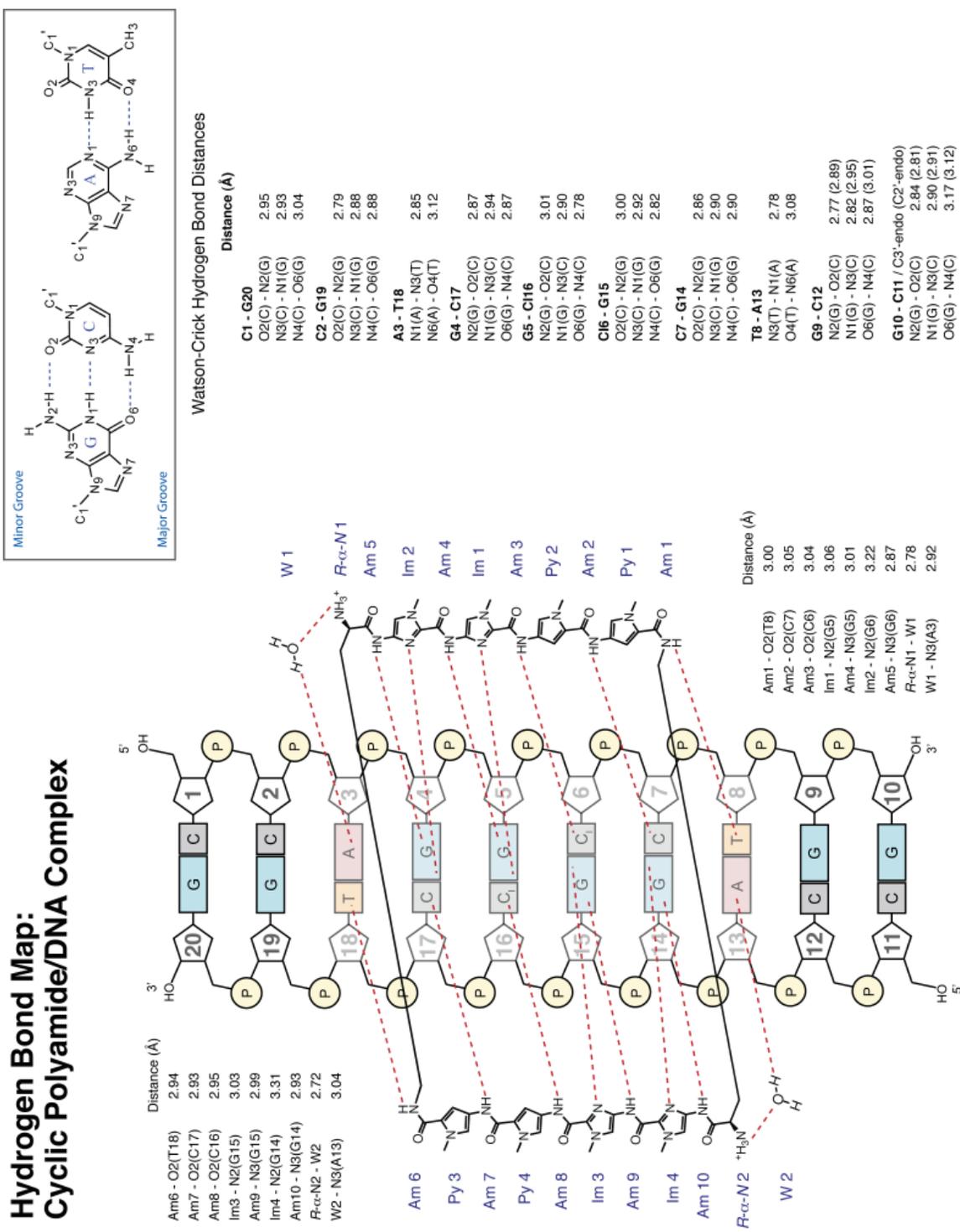
Data collection		DNA-Polyamide	DNA
	Space group	P1	C2
	Cell dimensions		
	a, b, and c, Å	22.50	31.83
		25.14	25.64
		29.09	34.17
	$\alpha$ , $\beta$ , and $\gamma$ , °	66.53	90
		79.28	116.72
		79.57	90
	Wavelength	0.970	0.999
	Resolution, Å	26.47 - 1.18	15.91 - 0.98
	$R_{\text{merge}}^*$	12.4 (15.4)	5.0 (7.5)
	$I/\sigma I^*$	12.5 (8.9)	18.7 (8.3)
	Completeness, %*	94.6 (90.7)	89.5
	Redundancy	4.0	2.5
Refinement			
	Resolution, Å	26.47 - 1.18	15.91 - 0.98
	No. of reflections	16,139	12,125
	$R_{\text{work}}/R_{\text{free}}^\ddagger$	9.9 / 13.8	10.8 / 14.4
	No. of atoms		
	DNA	441	297
	Polyamide	86	–
	Calcium	6	3
	Water	262	137
	B factors		
	DNA	5.6	6.5
	Polyamide	5.2	–
	Calcium	11.7	33.2
	Water	19.2	20.1
	R.m.s. deviations		
	Bond lengths, Å	0.026	0.033
	Bond angles, °	2.57	3.44

<sup>†</sup>Free  $R$  calculated against 5% of the reflections randomly removed.

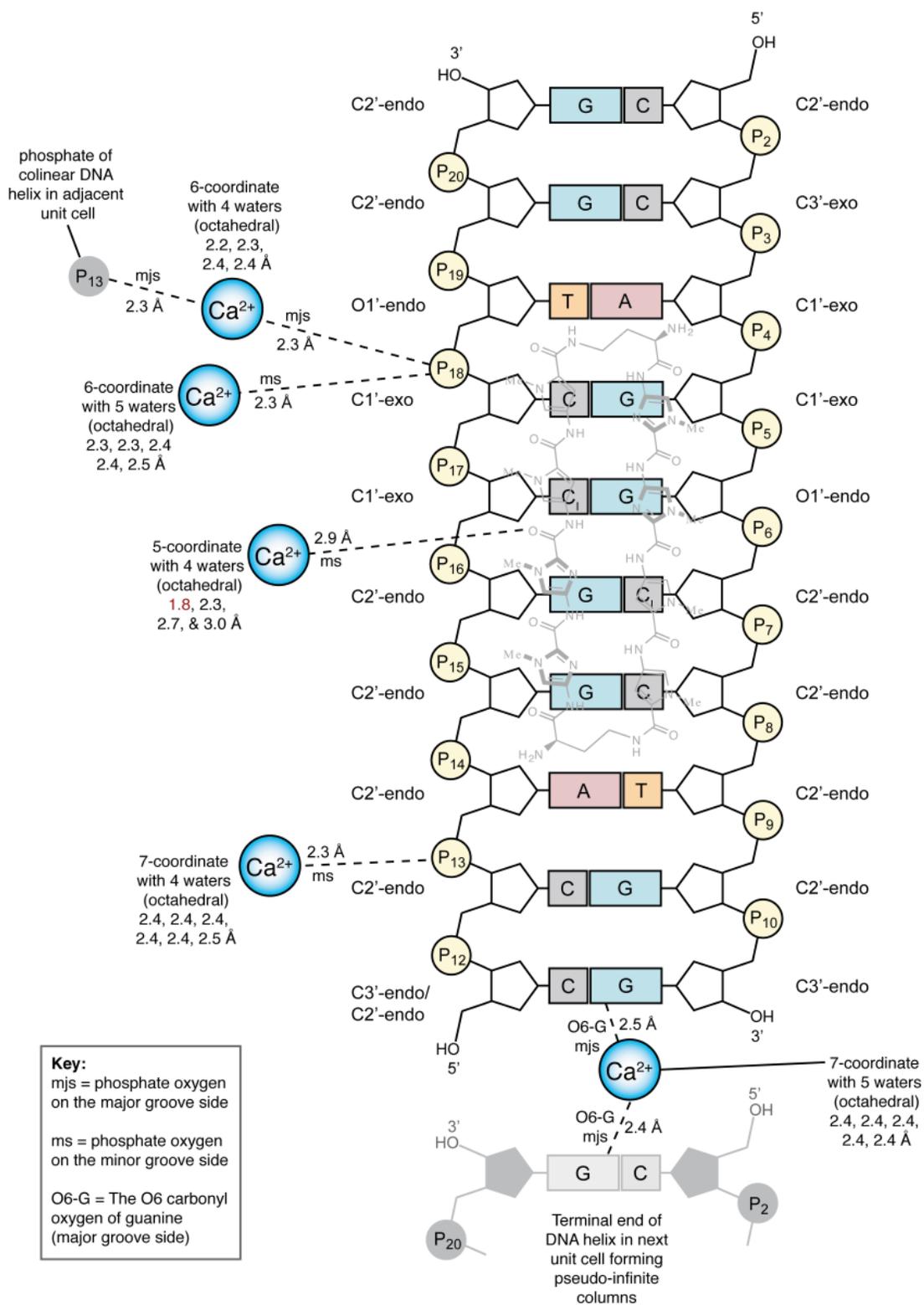
\*Highest-resolution shell is shown in parentheses.



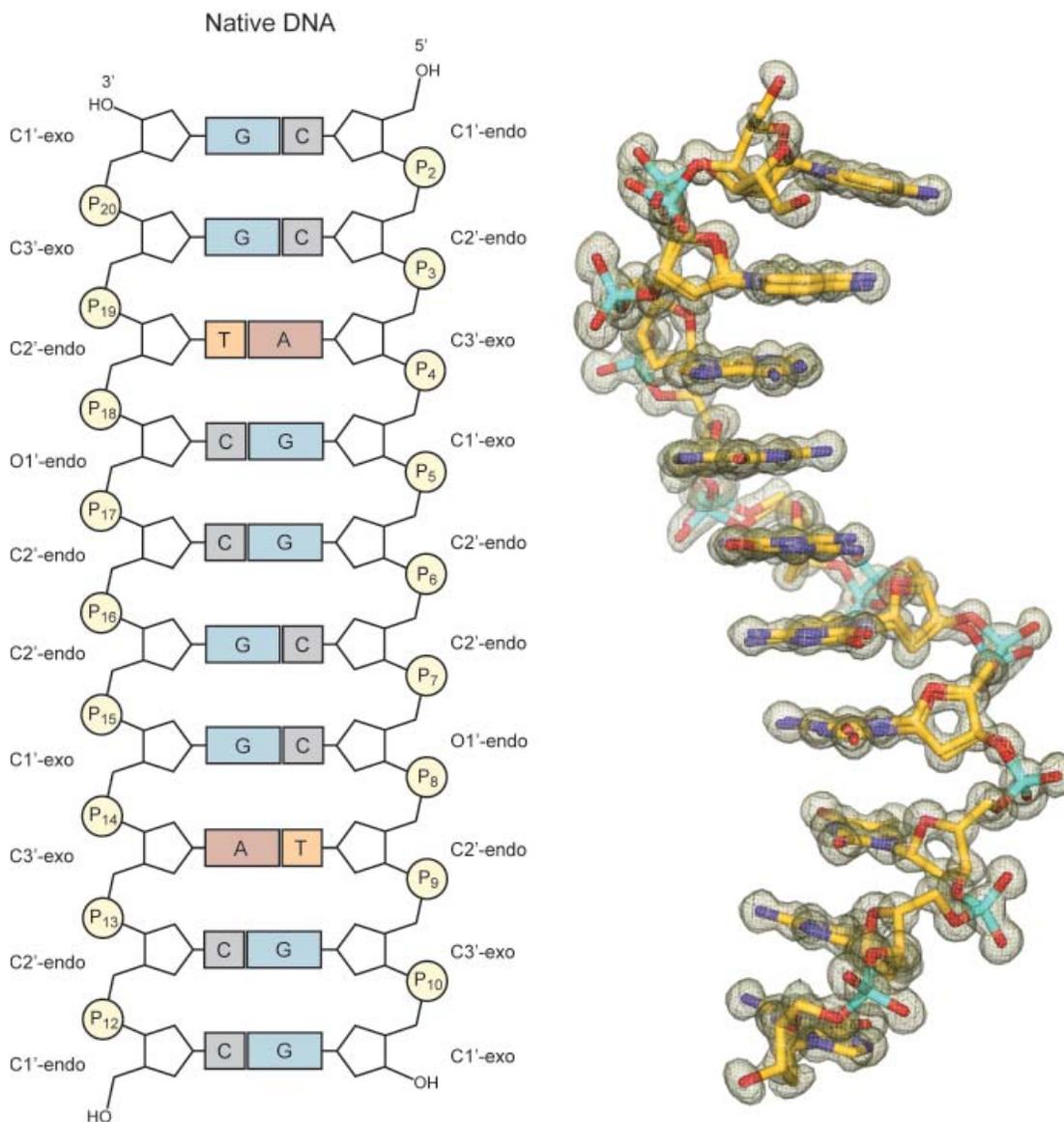
**Figure 5.8** Single crystal of the polyamide-DNA complex. a) Single crystal frozen in a cryoloop for data collection. b) UV-Vis of the single crystal polyamide-DNA complex dissolved in 50  $\mu\text{L}$  of water.



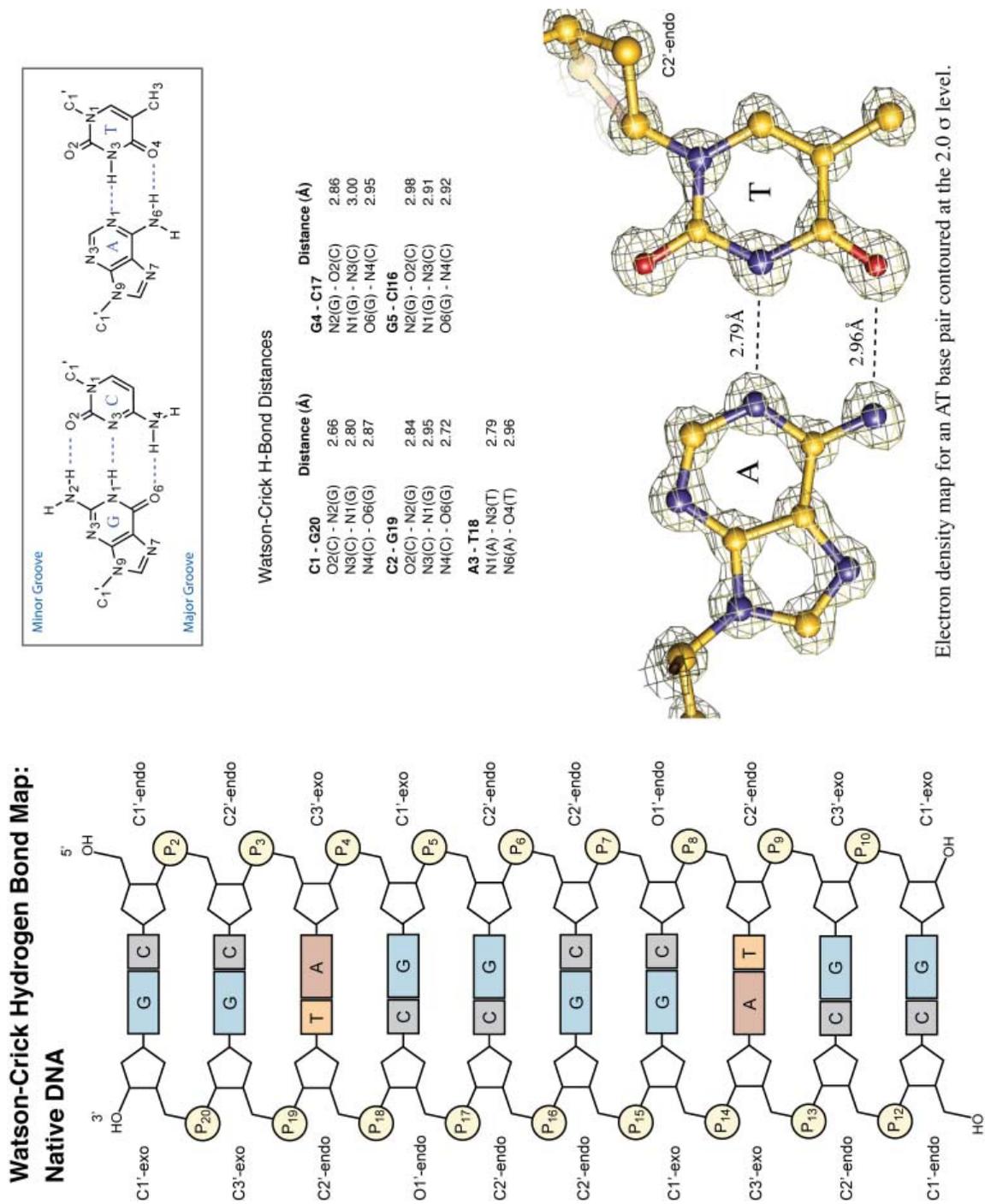
**Figure 5.9** Hydrogen bond map of polyamide-DNA complex.



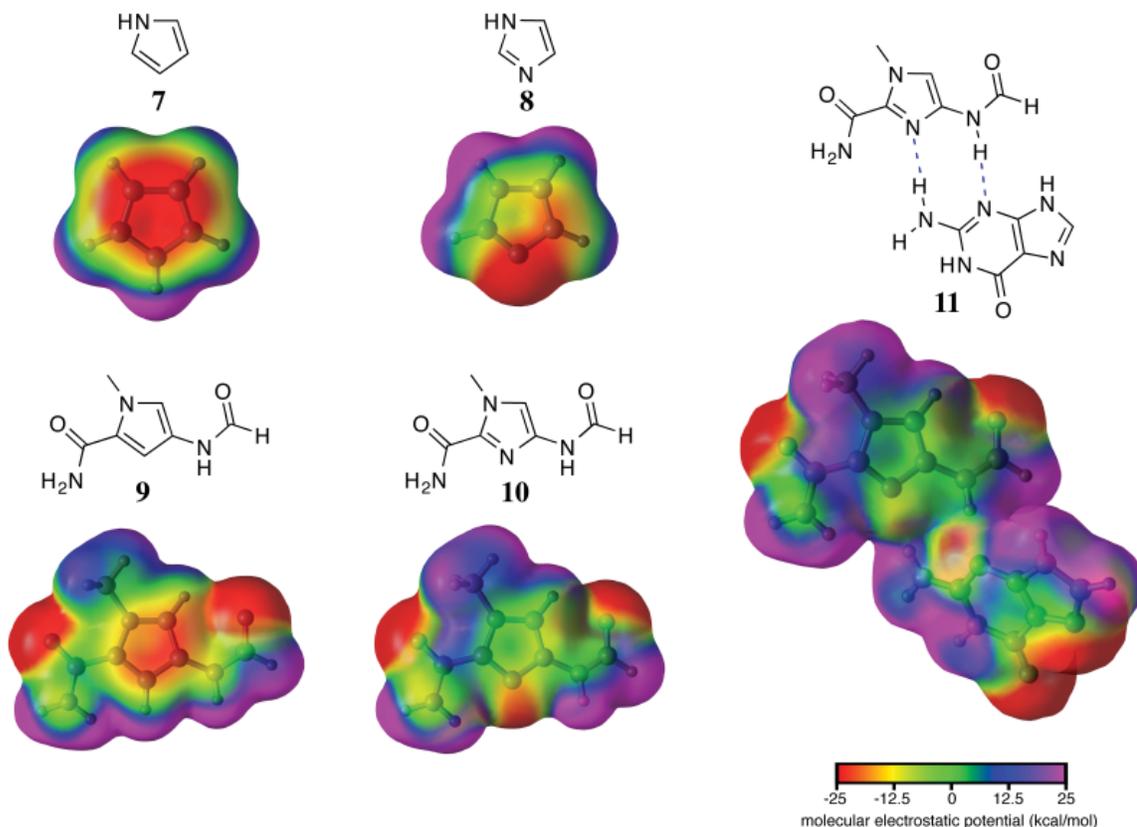
**Figure 5.10** Polyamide/DNA complex map. Schematic diagram of polyamide/DNA crystal structure showing the sugar conformation at each position and calcium ion coordination.



**Figure 5.11** DNA structure map. Schematic diagram of native DNA crystal structure showing the sugar conformation at each position and the electron density map of the DNA asymmetric unit contoured at the 1.0  $\sigma$  level. The structure on the right is oriented 5' to 3' from top to bottom and corresponds to the right hand strand in the structure map.



**Figure 5.12** Hydrogen bond map of DNA crystal structure and electron density of the AT base pair contoured at the 2.0  $\sigma$  level.

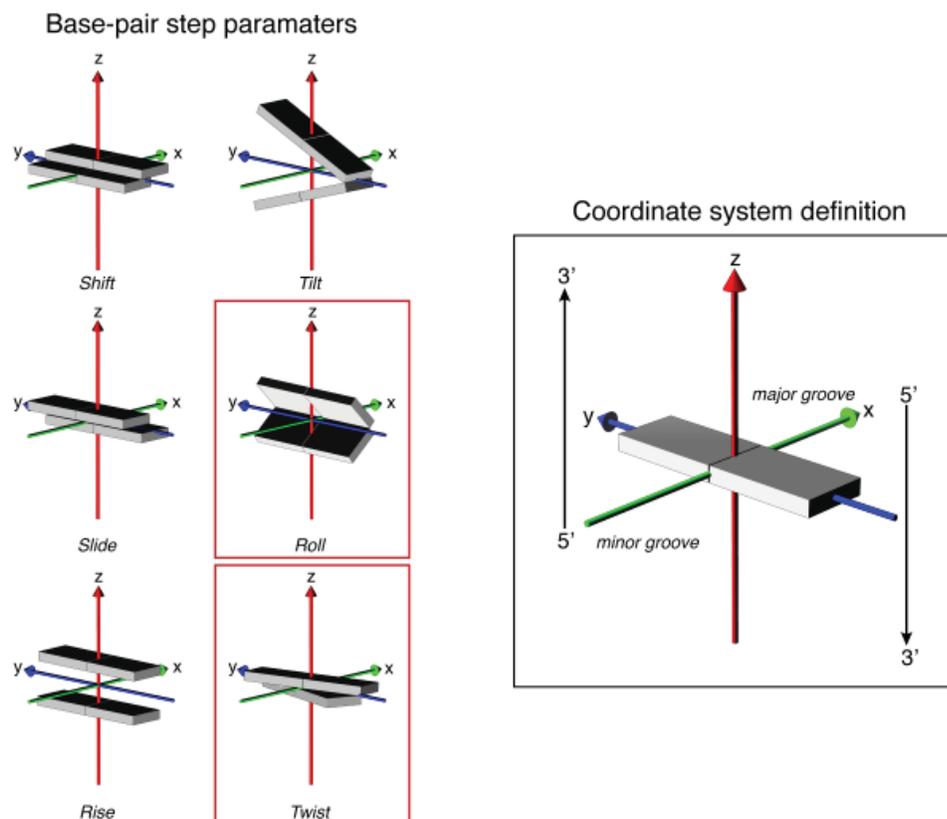


**Figure 5.13** Molecular electrostatic potential maps of compounds **7-10** and complex **11**. Complex **11** coordinates were taken directly from the polyamide/DNA crystal structure and used without geometry optimization. All ab initio calculations reported here were performed using HF/3-231G\* as implemented in the Gamess program.<sup>37-39</sup> Full geometry optimization was performed on all structures except for complex **11** whose coordinates correspond to those of the crystal structure. Electrostatic potential surfaces were generated by mapping the electrostatic potentials onto surfaces of molecular electron density ( $0.002 \text{ electron}/\text{\AA}^3$ ) and color-coding, using the Chimera program.<sup>40</sup> The molecular electrostatic potential energy values range from -25 kcal/mol for values of negative potential (red) to +25 kcal/mol for values of positive potential (blue). This range was chosen to emphasize the variations in the aromatic region and some regions of the electrostatic potential associated with heteroatoms may lie beyond the  $\pm 25$  kcal/mol range.

**Local base-pair step parameters**

Parameter*	CC/GG	CA/TG	AG/CT	GG/CC	GC/GC	CC/GG	CT/AG	TG/CA	GG/CC
Shift, Å									
DNA	-0.58	0.21	0.66	-0.34	0.00	0.34	-0.66	-0.21	0.58
PA/DNA	0.00	0.63	-1.96	-1.35	0.13	1.43	1.81	-1.78	0.09
Slide, Å									
DNA	0.69	2.80	0.78	0.24	-0.37	0.24	0.78	2.80	0.69
PA/DNA	-0.47	0.59	0.31	0.94	-0.02	0.84	0.38	1.51	0.29
Rise, Å									
DNA	3.30	3.26	3.08	2.99	4.48	2.99	3.08	3.26	3.30
PA/DNA	3.33	3.08	3.28	3.28	3.04	3.29	3.29	3.08	3.43
Tilt, °									
DNA	1.82	-0.81	-3.57	-5.49	0.00	5.49	3.57	0.81	-1.82
PA/DNA	4.14	1.42	-8.88	-4.81	0.64	5.05	5.38	-2.03	-3.89
Roll, °									
DNA	9.80	-8.73	9.85	-2.96	-4.11	-2.96	9.85	-8.73	9.80
PA/DNA	11.79	7.20	13.46	1.07	-0.20	2.57	16.09	6.54	11.05
Twist, °									
DNA	27.24	50.86	21.53	33.56	52.38	33.56	21.53	50.86	27.24
PA/DNA	36.26	29.92	33.52	38.71	33.84	38.32	31.32	37.22	38.20

\*Relationship between the bases composing the base pair.

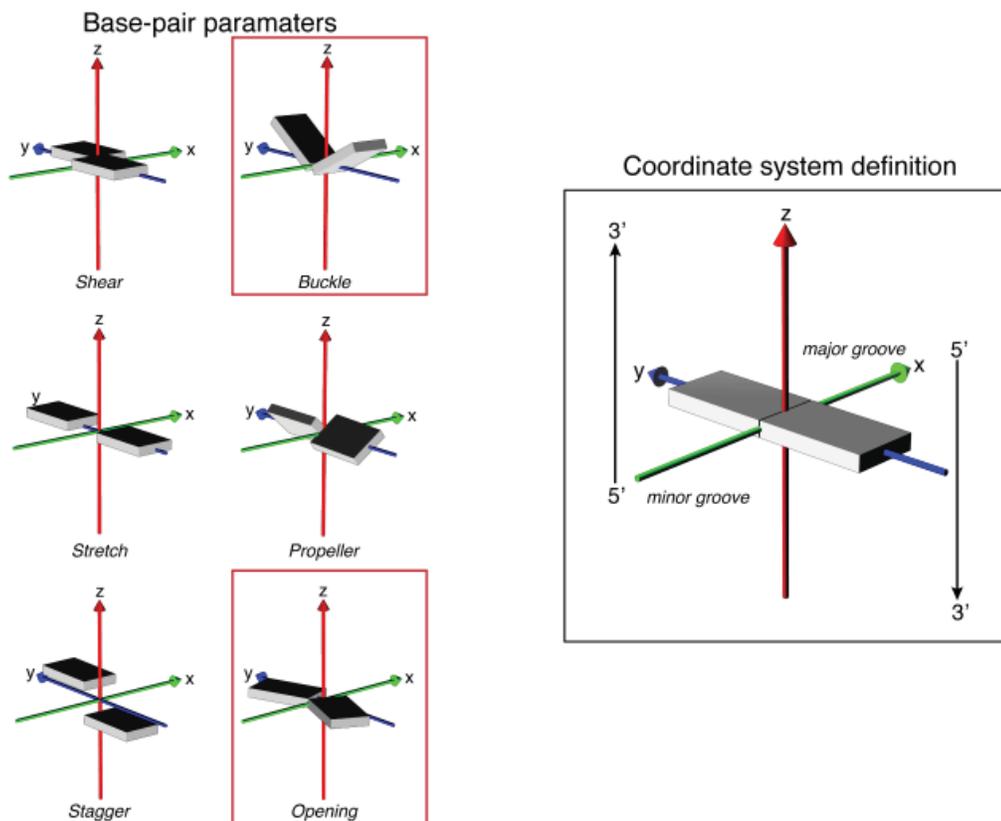


**Figure 5.14** Comparison of Local base-pair step parameters for DNA with and without polyamide complexed. Parameters were calculated using 3DNA.

## Local base-pair parameters and sugar pucker

Parameter*	C•G	C•G	A•T	G•C	G•C	C•G	C•G	T•A	G•C	G•C
Shear, Å										
DNA	0.04	0.16	0.02	-0.33	-0.32	0.32	0.33	-0.02	-0.16	-0.04
PA/DNA	0.15	0.10	0.16	-0.07	-0.17	0.11	0.10	-0.04	-0.37	0.05
Stretch, Å										
DNA	-0.21	-0.20	-0.13	-0.19	0.02	0.02	-0.19	-0.13	-0.20	-0.21
PA/DNA	-0.14	-0.12	-0.07	-0.10	-0.15	-0.13	-0.11	-0.13	-0.26	-0.09
Stagger, Å										
DNA	0.07	0.08	0.13	0.31	0.86	0.86	0.31	0.13	0.08	0.07
PA/DNA	0.16	0.09	-0.10	0.08	0.07	-0.02	-0.09	-0.21	0.03	0.17
Buckle, °										
DNA	-5.52	-2.64	0.23	13.52	22.43	-22.43	-13.52	-0.23	2.64	5.52
PA/DNA	-5.30	-9.09	5.62	-1.10	-2.90	1.60	1.81	0.42	11.29	5.28
Propeller, °										
DNA	-16.32	-9.30	-4.01	-9.51	-12.64	-12.64	-9.51	-4.01	-9.30	-16.32
PA/DNA	-21.09	-7.15	-12.99	-6.10	-16.87	-15.12	-1.60	-5.10	-9.33	-24.06
Opening, °										
DNA	2.09	-2.60	1.17	0.49	4.88	4.88	0.49	1.17	-2.60	2.09
PA/DNA	-0.69	-0.93	4.73	-1.94	-4.39	-4.48	-0.99	6.11	0.60	2.78
Sugar pucker										
DNA	C1'-exo C1'-exo	C2'-endo C3'-exo	C2'-endo C2'-endo	C1'-exo C1'-exo	C2'-endo C1'-exo	C1'-exo C2'-endo	C1'-exo C1'-exo	C2'-endo C2'-endo	C3'-exo C2'-endo	C1'-exo C1'-exo
PA/DNA	C3'-endo C2'-endo	C2'-endo C2'-endo	C2'-endo O4'-endo	C2'-endo C1'-exo	C2'-endo C1'-exo	C1'-exo C2'-endo	C1'-exo C2'-endo	C1'-exo C2'-endo	C3'-exo C2'-endo	C2'-endo C3'-endo

\*Relationship between the bases composing the base pair.

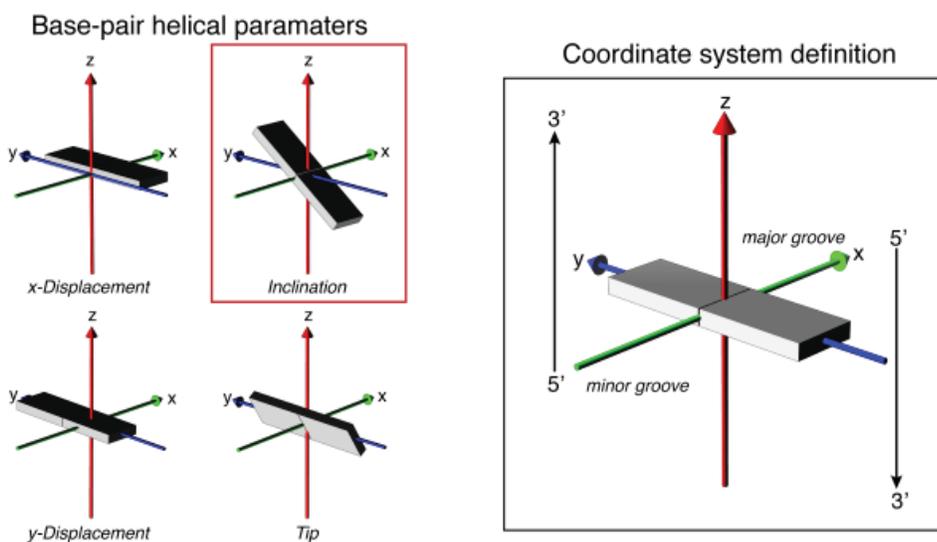


**Figure 5.15** Comparison of Local base-pair parameters and sugar conformations for DNA with and without polyamide complexed. Parameters were calculated using 3DNA.

**Local base-pair helical parameters**

Parameter*	CC/GG	CA/TG	AG/CT	GG/CC	GC/GC	CC/GG	CT/AG	TG/CA	GG/CC
X-displacement, Å									
DNA	-0.87	3.77	-1.26	0.83	-0.05	0.83	-1.26	3.77	-0.87
PA/DNA	-2.19	-0.26	-1.50	1.28	-0.01	0.94	-1.93	1.49	-0.93
Y-displacement, Å									
DNA	1.57	-0.30	-2.74	-0.22	0.00	0.22	2.74	0.30	-1.57
PA/DNA	0.52	-0.91	1.77	1.42	-0.12	-1.51	-2.12	2.48	-0.60
Inclination, °									
DNA	19.97	-10.08	24.57	-5.07	-4.65	-5.07	24.57	-10.08	19.97
PA/DNA	18.29	13.69	21.84	1.61	-0.34	3.89	27.43	10.14	16.42
Tip, °									
DNA	-3.71	0.94	8.91	9.41	0.00	-9.41	-8.91	-0.94	3.71
PA/DNA	-6.42	-2.70	14.41	7.22	-1.09	-7.64	-9.18	3.15	5.78

\*Relationship between the bases composing the base pair.



**Figure 5.16** Comparison of Local base-pair helical parameters for DNA with and without polyamide complexed. Parameters were calculated using 3DNA.

## **Chapter 6: Structural Elucidation of a $\beta$ -amino- $\gamma$ -linked Cyclic**

### **Polyamide-DNA Complex and RNA Binding Studies**

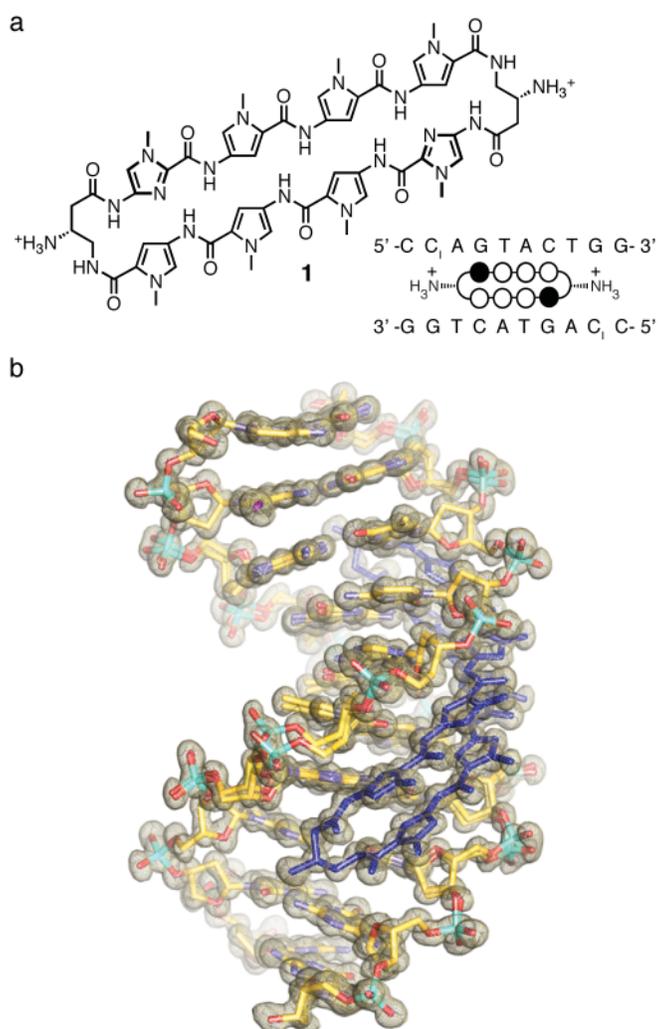
**Abstract**

Chemical control of dysregulated gene expression poses a significant challenge at the frontier of chemical biology. Py-Im polyamides are a class of small molecules that can be programmed to bind a broad repertoire of DNA sequences, disrupt transcription factor-DNA interfaces, and modulate gene expression pathways in cell culture experiments. Detailed structural information thus far has been limited to moderate resolution X-ray structures of unlinked 2:1 binding polyamides and NMR NOESY-restrained molecular dynamics models, with atomic resolution X-ray structures remaining elusive. Structural elucidation of polyamide-DNA complexes is fundamental to understanding the recognition properties at the molecular level and this chapter reports a high resolution (0.95 Å) X-ray crystal structure of a cyclic Py-Im polyamide bound to the central six base pairs of the sequence d(5'-CCAGTACTGG-3')<sub>2</sub>, revealing the hydrophobic, electrostatic, and shape selective recognition interactions responsible for DNA binding. Additionally, a structural basis for the allosteric modulation of transcription-factor DNA interfaces with β-amino turn linked cyclic polyamides is reported and a combination of biophysical, structural, and modeling studies are presented which explain the inability of polyamides to bind double helical A-form RNA.

## 6.1 Introduction

Pyrrole-imidazole polyamides are a class of small molecules that bind the minor groove of DNA sequence specifically.<sup>1,2</sup> Encoded by side-by-side arrangements of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamide monomers, Im/Py pairs distinguish G•C from C•G base pairs, whereas Py/Py pairs are degenerate for T•A and A•T.<sup>3-6</sup> Polyamides have been shown to bind a broad repertoire of DNA sequences,<sup>7</sup> permeate cell membranes and traffic to the nucleus,<sup>8-15</sup> access chromatin,<sup>16,17</sup> and disrupt protein-DNA interactions.<sup>2</sup> Polyamide inhibition of transcription factor-DNA binding of HIF-1 $\alpha$ ,<sup>18,19,20</sup> androgen receptor (AR),<sup>21</sup> and AP-1<sup>22,23</sup> has been exploited for controlling expression of medically relevant genes such as VEGF, PSA, TGF- $\beta$ 1 and LOX-1 in cell culture experiments. X-ray crystallography has provided structures of unlinked 2:1 binding polyamides to a resolution of 2.00 Å, providing valuable insight into the polyamide-DNA molecular recognition process.<sup>5,6,24</sup> Structural studies of hairpin polyamides bound to the nucleosome core particle have also provided structural proof that polyamides can bind biologically relevant higher order structure, however a combination of resolution limits and high B-factors prevented a detailed picture of the polyamide-DNA interactions beyond confirmation of the binding location.<sup>16,25</sup> Much insight has been gleaned from NMR studies where NOESY-restrained molecular dynamics models have provided structures of 1:1 and 6-ring cyclic polyamides.<sup>26,27</sup> Despite these successes, atomic resolution X-ray structures of this important class of compounds and in general minor groove binders have remained elusive.

The ability of DNA to undergo bending, twisting, and stretching motions as well as the long-range propagation (allosteric effect) of these perturbations coupled with DNA recognition by proteins and small molecules can have profound influences over important processes such as gene transcription and modulation of eukaryotic gene networks.<sup>28-32</sup> Allosteric communication along and through the DNA helix has been the subject of intense study and forms the basis for cooperative interactions among transcription factor regulatory networks such as the interferon- $\beta$  enhanceosome, where transcription factor binding induced DNA conformational changes led to cooperative enhancer occupancy. This potential for short and long range allosteric control over the transcriptional machinery provides a powerful concept for the design of small molecules that can bind to topographically distinct locations on DNA with the possibility of modulating transcription factor-DNA binding by allosteric perturbation of the DNA structure.<sup>30,31</sup> Recent studies (Chapter 3) of cyclic polyamide **1** and analogous hairpin polyamides revealed they possessed high DNA binding affinities and could regulate endogenous androgen receptor-activated gene expression (prostate



**Figure 6.1** Structure of cyclic polyamide **1** presented with its ball-and-stick model superimposed over the binding site on the dsDNA oligonucleotide sequence used for crystallization. Closed circles designate *N*-methylimidazole, open circles designate *N*-methylpyrrole, and half open circles substituted with ammoniums designate  $\beta$ -amino substituted  $\gamma$ -turn unit. b. Electron density map contoured at the 1.0  $\sigma$  level for the X-ray crystal structure of cyclic polyamide **1** complexed to dsDNA (0.95 Å resolution).

specific antigen) in cell culture, from which we infer cell permeability. Additionally, In vitro ADMET studies of cyclic **1** and hairpin polyamides revealed favorable drug-like properties for both classes of compounds and excellent metabolic stability.

Reported, is the atomic resolution structure (0.95 Å resolution) of an 8-ring cyclic polyamide in complex with double helical DNA. The cyclic polyamide **1** is comprised of two antiparallel ImPyPyPy strands capped by (*R*)- $\beta$ -amino- $\gamma$  turn units. Polyamide **1**, which codes for the sequence 5'-WGWWCW-3' was cocrystallized with the palindromic DNA oligonucleotide sequence 5'-CC<sub>1</sub>AGTACTGG-3' 10 base pairs in length (Fig. 6.1). We observe significant allosteric structural perturbations of the DNA helix induced upon binding of substituted GABA ( $\gamma$ -aminobutyric acid) turn-linked polyamides in the DNA minor groove. In addition to amide and imidazole recognition with the DNA minor groove floor, a detailed view of the  $\beta$ -amino- $\gamma$ -turn conformation and hydration reveals a network of well-

ordered water-mediated interactions between the polyamide and DNA. Significantly, we find that a conformational inversion occurs at the turn position upon moving the amino substituent from the  $\alpha$  to  $\beta$  positions of the GABA turn. The allosteric modulation of the DNA structure induced by polyamide binding is also shown and a structural basis for the inability of polyamides to bind A-form

**Table 6.1** Data collection and refinement statistics.

Data collection		
Space group		P4 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions		
a, b, and c, Å		39.83 39.83 84.57
$\alpha$ , $\beta$ , and $\gamma$ , °		90 90 90
Wavelength		0.82654
Resolution, Å		23.44 - 0.95
$R_{\text{merge}}^*$		9.0 (52.7)
$I/\sigma I^*$		13.9 (2.3)
Completeness, %*		97.7 (99.6)
Redundancy		7.4
Refinement		
Resolution, Å		23.44 - 0.95
No. of reflections	39,782	
$R_{\text{work}}/R_{\text{free}}^\ddagger$		11.2 / 12.4
No. of atoms		
DNA		445
Polyamide		86
Calcium		4
Water		239
B factors		
DNA		7.2
Polyamide		6.7
Calcium		13.2
Water		23.1

\*Highest-resolution shell is shown in parentheses.

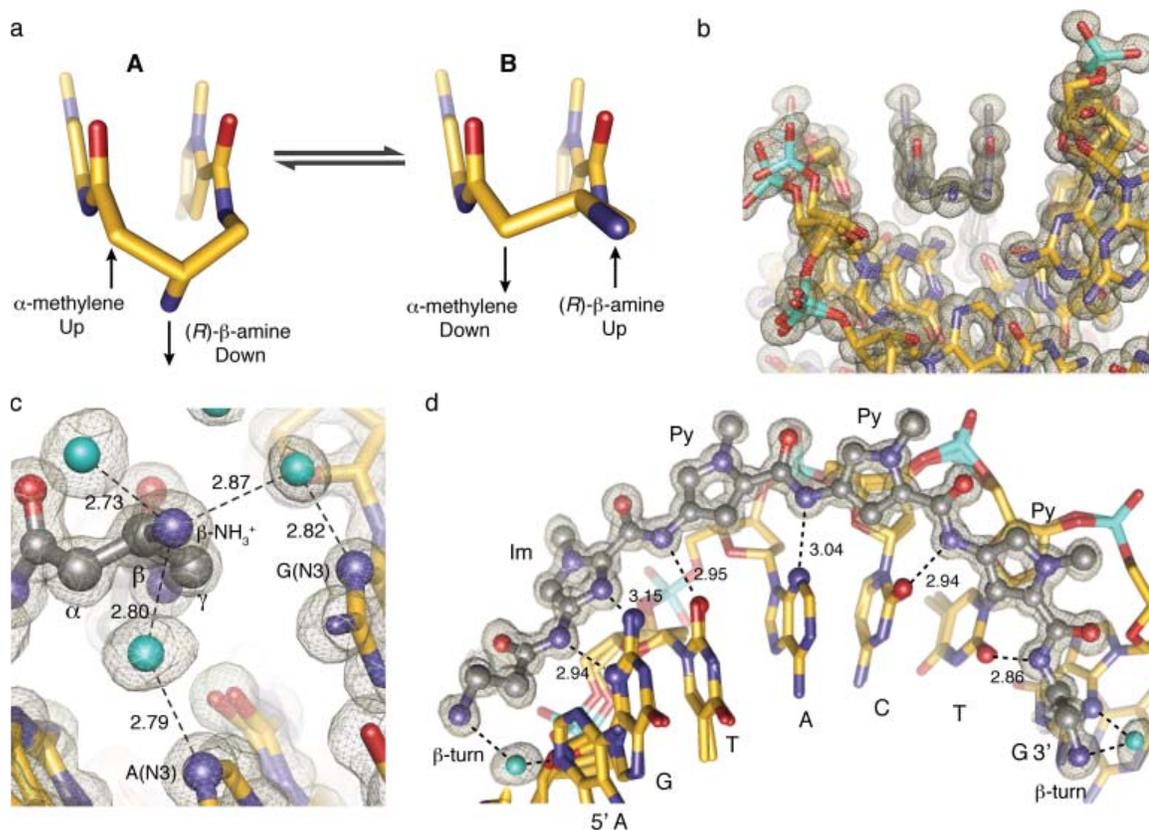
†Free  $R$  calculated against 5% of the reflections randomly removed.

RNA is presented with UV-melting temperature data.

## 6.2 Overall Structure

The structure of cyclic polyamide **1** in complex with d(5'-CC<sub>1</sub>AGTACTGG)<sub>2</sub> was solved by direct methods to 0.95 Å resolution with synchrotron radiation (Fig. 6.1). One cyclic polyamide bound to a single DNA duplex is present in the asymmetric unit of the crystal in the P4<sub>1</sub>2<sub>1</sub>2 space group. The final structure was refined anisotropic and unrestrained to an  $R$ -factor of 11.2% and an  $R_{\text{free}}$  of 12.4% (Table 6.1). The average B-factor for the polyamide was 6.7 Å<sup>2</sup> and 7.2 Å<sup>2</sup> for DNA. The asymmetric unit contains one full polyamide-complexed DNA double helix. In the DNA complex, the aromatic amino acids are bound with an N- to C-orientation of each ImPyPyPy strand of

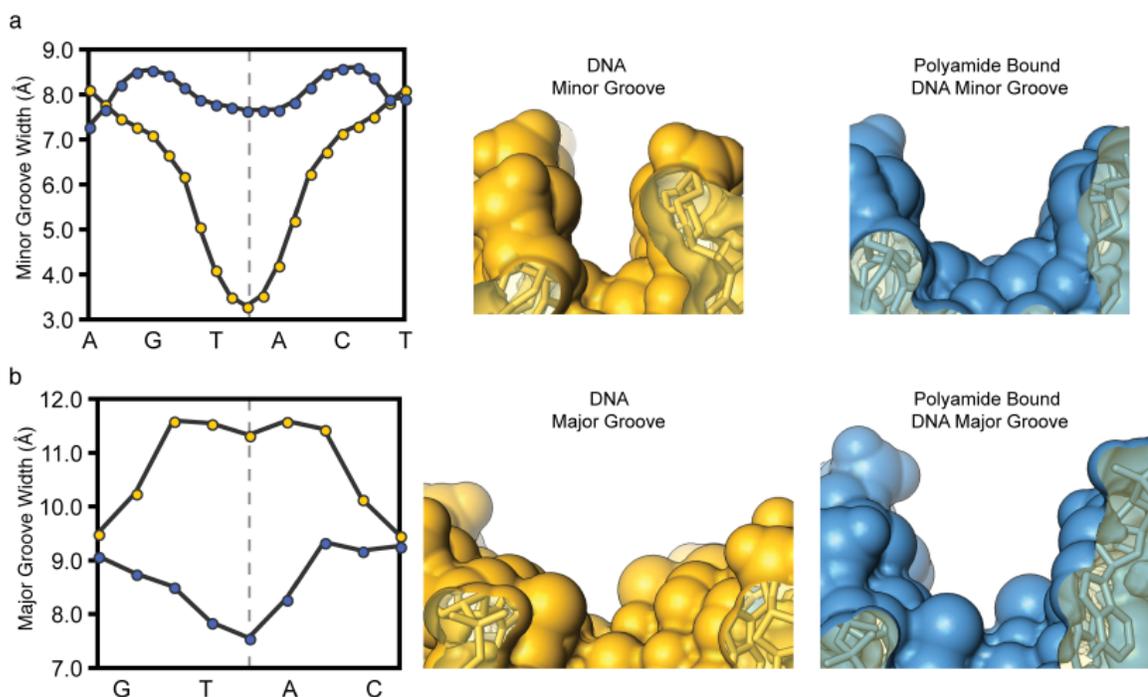
the cycle adjacent to the 5' to 3' direction of the DNA. The conformational constraints imposed by the turn unit result in ring placement that is ring-over-ring as opposed to ring-over-amide. Greater than 40% of the polyamide surface area is buried leaving only the top of the methyl groups on the heterocycles, the amide carbonyl oxygens, and the chiral  $\beta$ -ammonium turn solvent exposed. Alternate conformations are observed for 7 of the 18 nucleotides of the DNA duplex. The cyclic pyrrole-imidazole polyamide **1** was co-crystallized with the palindromic DNA oligonucleotide sequence shown in Figure 6.1. The polyamide selectively binds to the sequence, 5'-WGTACW-3', and previous studies have demonstrated that the equilibrium binding constants are sub-nanomolar and outside the measurement range of DNase I footprinting methods (see Chapter 3).



**Figure 6.2** Molecular recognition details from the X-ray structure of cyclic polyamide **1**. a. Conformation of the  $\beta$ -amino substituted GABA turn linkage. Conformation A (left) is the conformation observed in the previously determined  $\alpha$ -amino turn X-ray crystal structure. Conformation B (right) shows the preferred conformation for the  $\beta$ -amino turn determined by X-ray crystallography in this report. The  $\beta$ -methine conformational preference is puckered up and away from the DNA minor-groove floor, aligning the  $\beta$ -ammonium along the groove floor. b. Structural view looking down the DNA minor groove showing the bound cyclic polyamide with electron density contoured at the  $1.0 \sigma$  level. c. Geometry of the alpha-amino turn interacting with the adenine and guanine base pairs in the floor of the DNA minor groove through water-mediated hydrogen bonds. d. Isolated view of one half of the polyamide (split along a plane through the long axis of the polyamide and the DNA helical axis) showing hydrogen bond distances made to the DNA minor groove floor. Hydrogen bonding interactions of the DNA-polyamide complex with electron density contoured at the  $1.0 \sigma$  level. (Im = imidazole and Py = pyrrole)

### 6.3 Overall structure of DNA-polyamide complex

The cyclic pyrrole-imidazole polyamide is bound in an antiparallel head-to-tail turn-linked fashion with the N- to C-terminal orientation of each PyPyPyIm strand of the polyamide directly adjacent to a DNA strand oriented in a 5' to 3' direction. Figure 6.1 shows the overall structure of the complex with the electron density map contoured to the  $1.0 \sigma$  level. Figure 6.2b shows a view of the complex looking directly down the minor groove at the polyamide turn linkage. From this view it can be seen that significant stabilization of the complex is derived from van der Waals



**Figure 6.3** DNA minor and major groove dimensions in the absence and presence of polyamide. Native DNA structure  $d(5'-CCAGTACTGG-3')_2$  solved by Rees and coworkers (PDB: 1d8g, 0.74 Å resolution). a. Comparison of the minor groove width for DNA in the absence of polyamide (yellow curve and structure) and in the presence of bound polyamide (blue curve and structure). b. Comparison of the major groove width for DNA in the absence of polyamide (yellow curve and structure) and in the presence of bound polyamide (blue curve and structure). note: polyamide has been removed from the blue complex for clarity.

interactions between the outside face of the pyrrole-imidazole heterocyclic strands and the walls of the minor groove, which form a deep binding pocket for the polyamide. Greater than 40% of the polyamide surface area is buried and not solvent exposed, leaving only the top of the methyl groups of the heterocycles, the amide carbonyl oxygens, and the chiral  $\beta$ -amino turn linkage solvated. The turn linkage adds a conformational constraint to the ends of the polyamides preventing the heterocycle strands from slipping past each other as observed in the slipped orientations found in some 2:1 binders such as distamycin.

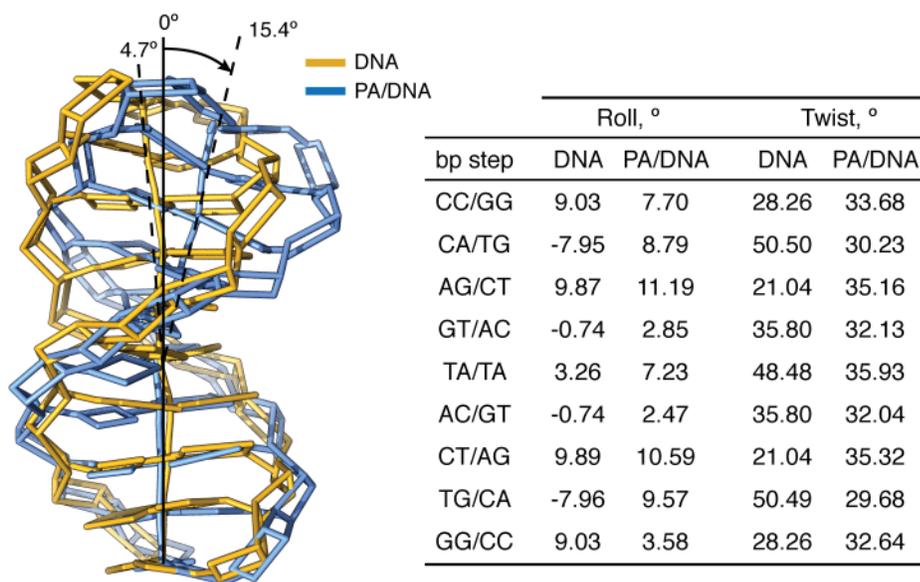
#### 6.4 Turn conformation

The  $\beta$ -methine conformational preference is puckered up and away from the DNA minor-groove floor, aligning the  $\beta$ -ammonium along the groove. Conformation A (left) is the conformation observed in the previously determined  $\alpha$ -amino turn X-ray crystal structure. Conformation B (right) shows the preferred conformation for the  $\beta$ -amino turn determined by

X-ray crystallography in this report. The hydration pattern around the turn is highly conserved at both ends of the structure and there are two water-mediated hydrogen bonds within 2.79 - 2.87 Å from the ammonium to the DNA minor-groove floor (Fig. 6.2c). The amide NH's and imidazole lone-pairs form a continuous series of direct hydrogen bonds to the floor of the DNA minor-groove, while the imidazoles impart specificity for the exocyclic amine of guanine through relief of steric interaction and a G(N2-hydrogen)-Im (lone pair) hydrogen bond (Fig. 6.2c-d). The amides linking the aromatic rings and the turns contribute hydrogen bonds to the purine N3 and pyrimidine O2 lone pairs where they all are within hydrogen bonding distance of a single DNA base. In total there are 10 direct amide hydrogen bonds (2.86–3.08 Å), 2 direct imidazole hydrogen bonds (3.15 and 3.16 Å), and 4 (*R*)-β-ammonium turn water-mediated hydrogen bonds (two per turn, average distance from amine to water = 2.79 Å–2.87 Å) to the floor of the DNA minor groove with at least one interaction for all 12 DNA base-pairs in the 6 bp binding site. There are a total of 16 hydrogen bond interactions between the cyclic polyamide and the floor of the DNA minor-groove, utilizing every hydrogen-bond donor and acceptor of the ligand (Fig. 6.2).

### 6.5 Allosteric Perturbations

Polyamide binding induces large structural changes in DNA and Figure 6.3 shows a



**Figure 6.4** Allosteric distortion upon polyamide binding. (left) DNA bending is observed for polyamide-bound DNA (blue structure) versus unbound-DNA (yellow structure). (right) Helical parameters for DNA in the absence and presence of polyamide showing an increase in positive roll and significant changes in twist angles upon polyamide binding. (Polyamide has been removed from the blue complex for clarity)

**Table 6.2** Buckle and opening values.

bp step	Buckle, °		Opening, °	
	DNA	PA/DNA	DNA	PA/DNA
C•G	-6.41	-10.59	0.99	0.68
C•G	-0.43	-7.08	-1.15	1.71
A•T	-1.63	6.15	1.35	3.84
G•C	10.75	-2.23	1.26	-0.78
T•A	13.15	0.37	-0.27	-4.87
A•T	-13.15	1.17	-0.27	-4.88
C•G	-10.74	2.39	1.27	-1.10
T•A	1.63	-2.93	1.34	3.78
G•C	0.44	16.58	-1.13	0.87
G•C	6.42	23.00	1.00	-0.15

slice through the short axis of the DNA helix showing the minor and major groove geometry at the center of the polyamide binding site for uncomplexed and complexed DNA. Figure 6.3 reveals a  $>4$  Å widening of the DNA minor groove upon polyamide binding and a compression of the major groove by more than 4 Å. Additionally, Figure 6.3 shows a major perturbation in the major groove depth upon polyamide binding converting the wide shallow surface of the major groove from a functionally exposed protein recognition domain to a narrow deep cleft less likely to accommodate the width of a standard protein alpha-helical domain or beta-sheet from a transcription factor. Figure 6.4 shows the polyamide induced bending of the DNA helix. The helix is bent toward the major groove by  $>15^\circ$  resulting in major groove compression. The base-pair step parameters in Figure 6.4 show a large positive roll throughout the polyamide binding site which contributes to the significant bend in the DNA helix. Additionally, polyamide binding induces a more uniform helical twist resulting in less variability as the base-pair step changes. The helical twist values for polyamide bound DNA range from  $29.68$ - $35.93^\circ$ . Values for the helical twist are highly sequence dependent in native DNA and range from  $21.04$  to  $50.50^\circ$  depending on step sequence. Major perturbations in the DNA base pair buckle and opening are also observed upon polyamide binding. At the central 4 base pairs of the binding site the buckle is significantly reduced upon binding and the base pairs are opened toward the DNA major groove with the largest variations occurring at the central AT base pairs. [Note: For a full set of helical parameters and definitions see Figures 6.6, 6.7, and 6.8.]

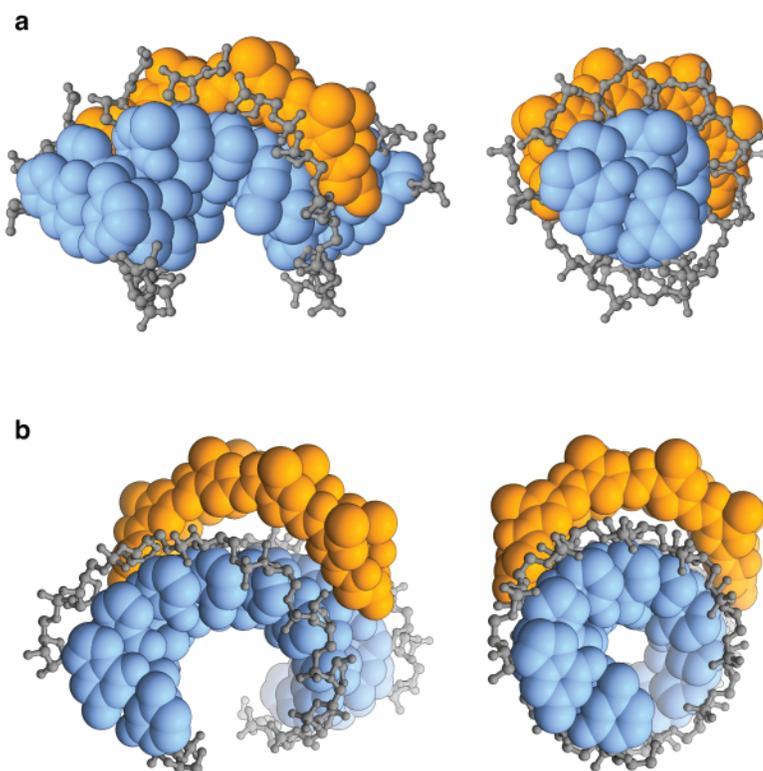
## 6.6 Solvation

The structure has a cell volume of  $134,162 \text{ \AA}^3$  and a Matthews Coefficient of 2.24 with a solvent content of 51%. There are 130 out of 239 water molecules within  $3.0 \text{ \AA}$  of the polyamide-DNA complex and 76 of the 130 water molecules localized around the DNA phosphates. The solvent exposed surface of the polyamide is hydrated by 22 of the 130 waters found within  $3.0 \text{ \AA}$  of the complex.

**Table 6.3** Polyamide-DNA and Polyamide-RNA melting temperatures.

Polyamides	dsDNA sequence		dsRNA sequence	
	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$
—	46.1 ( $\pm 0.8$ )	—	60.4 ( $\pm 0.6$ )	—
 <b>(1)</b>	82.8 ( $\pm 0.6$ )	36.6 ( $\pm 1.0$ )	59.9 ( $\pm 0.8$ )	-0.5 ( $\pm 1.0$ )
 <b>(2)</b>	72.3 ( $\pm 0.5$ )	26.2 ( $\pm 1.0$ )	60.7 ( $\pm 0.5$ )	0.3 ( $\pm 0.8$ )
 <b>(3)</b>	52.8 ( $\pm 0.4$ )	6.6 ( $\pm 0.9$ )	59.7 ( $\pm 0.3$ )	-0.6 ( $\pm 0.7$ )

These waters cluster and form bridges across the carbonyl oxygens of adjacent amides linking ring pairs, resulting in stripes of well ordered water across the polyamide surface. Additionally, 6 waters hydrate the polyamide ammonium turns (3 at each turn) with 4 of the 6 anchoring the polyamide to the floor of the DNA minor groove through bridging hydrogen bonds.



**Figure 6.5** a. Crystal structure of DNA-polyamide complex showing shape complementary and favorable hydrophobic interactions with the sugar-phosphate backbone. b. Coordinates of the cyclic polyamide docked within van der Waals radius of the putative binding site on a model of ideal A-form double helical RNA.

## 6.7 RNA Binding Studies

The minor groove binding of polyamides to B-form DNA has been extensively studied, however the ability of polyamides to bind double helical RNA has received little attention. There are two major differences in helical RNA versus DNA. First, thymine (T) is replaced by uracil (U) presenting the addition of a 5' methyl group to the major groove of the helix. However, the hydrogen

bonding functionality of the minor groove remains identical to that of B-form DNA. The second and most important difference is the addition of a 2'-OH on the sugar resulting in ribose as opposed to deoxyribose for DNA. This extra hydroxyl has a profound effect on the overall helical RNA structure and rigidity primarily due to the enforcement of a C3'-endo ribose sugar pucker. This pucker forces the RNA helix into an A or A'-form conformation due to the steric incompatibility of the 2'-OH with a DNA B-form conformation, which prefers a C2'-endo sugar conformation. The conformational rigidity leads to less sequence dependent microstructure than DNA and a dramatically different minor and major groove geometry. Additionally, the DNA helix has been shown to be highly conformationally mobile in contrast to the RNA helix. The structure of A-form RNA has an 11-fold helix with a narrow, deep major groove and shallow, wide minor groove in stark contrast to B-form DNA. The base pairs of A-form RNA are inclined and drastically displaced from the helix axis causing an overall expansion of the helix width, which in turn leads to a dramatically shallow curvature of the minor groove floor. The criteria required for polyamide binding and, in general, small molecule binding relies on the minimization of water exposed hydrophobic surfaces, the complementary pairing of buried hydrogen bond donors and acceptors, the maximization of van der Waals interactions, the solvation or neutralization of all charges, and the maximization of attractive and minimization of repulsive interactions. Our results show that the polyamides in this study provide a large thermal stabilization to DNA as opposed to RNA, which does not have an increased melting temperature for any of the compounds studied. The thermal stabilization for the DNA duplex ranges from 4 °C for the unlinked polyamide (2:1 complex at saturating concentrations) to 23 °C for the cyclic polyamide.

## 6.8 Conclusion

The crystal structure presented highlights the molecular recognition of  $\beta$ -amino turn-linked polyamides in the minor-groove of DNA and provides insight into the allosteric modulation of B-form DNA by Py-Im polyamides. The DNA structural distortion induced upon polyamide minor-groove binding provides an allosteric model for disrupting DNA:transcription factor interfaces in the promoters of selected genes. The ability of DNA to undergo short and long-range allosteric effects coupled with DNA binding by proteins can have influence over important processes such as modulation of eukaryotic gene networks. The potential for allosteric control over the transcriptional machinery provides a powerful concept for the design of small molecules that can bind to distinct locations on DNA with the possibility of modulating transcription factor activity. The RNA binding

studies demonstrate that this class of Py-Im polyamides are completely selective for binding dsDNA over dsRNA.

## 6.9 Experimental

### 6.9.1 Synthesis

Polyamides **1–3** were synthesized by standard solution-phase synthesis methods presented in Chapter 2 and 3 of this thesis. Synthetic deoxyoligonucleotides were purchased HPLC purified from Trilink Biotechnologies and desalted using a 5 gram sep-pak C18 cartridge (Waters) followed by lyophilization to dryness.

### 6.9.2 Oligonucleotide purification and Crystallization

Oligonucleotides were purchased HPLC purified from Trilink Biotechnologies (San Diego, CA). Prior to use the oligonucleotides were de-salted using a 5 gram sep-pak C18 cartridge (Waters) and lyophilized to dryness.

Single stranded DNA was quantitated by UV-Vis spectroscopy and incubated with a 2:1 ratio of DNA to polyamide prior to crystallization. Crystals were obtained from a solution of 0.6 mM duplex DNA, 0.75 mM polyamide, 24% 2-methyl-2,4-pentanediol (MPD), 35 mM calcium acetate, 10 mM Tris pH 7.5 equilibrated in sitting drops against a reservoir of 35% MPD at 4°C and crystals were flash cooled at 100 K prior to data collection. DNA crystals grew in space group  $P4_12_12$  with unit cell dimensions  $a = 39.8270\text{\AA}$ ,  $b = 39.8270\text{\AA}$ ,  $c = 84.5718\text{\AA}$ ,  $\alpha = 90.00^\circ$ ,  $\beta = 90.00^\circ$ ,  $\gamma = 90.00^\circ$ .

### 6.9.3 Data collection, Structure determination, and refinement

Polyamide-DNA crystals grew in space group  $P4_12_12$  with unit cell dimensions  $a = 39.8270\text{\AA}$ ,  $b = 39.8270\text{\AA}$ ,  $c = 84.5718\text{\AA}$ ,  $\alpha = 90.00^\circ$ ,  $\beta = 90.00^\circ$ ,  $\gamma = 90.00^\circ$ , and one polyamide-duplex DNA complex in the asymmetric unit. This data set was collected at Stanford Synchrotron Radiation Laboratory (SSRL) beamline 12-2 with a MAR Research imaging plate detector at wavelength 0.97 Å.

Data was processed with Mosflm<sup>33</sup> and scaled with the CCP4<sup>34</sup> program suite. Solution of both structures was obtained using SHELXD<sup>35,36</sup> direct methods. All atoms were visible in initial maps from direct methods solution. Refinement was performed using Refmac<sup>37</sup> and model building using Coot.<sup>38</sup>

#### 6.9.4 Structure Analysis

DNA helical parameters were calculated using the program Curves<sup>39</sup> and 3DNA.<sup>40</sup> Figures were prepared and measurements made using UCSF Chimera.<sup>41</sup>

#### 6.10 Notes and References

1. Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.
2. Dervan, P. B., and Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.
3. Trauger, J. W., Baird, E. E., and Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature.* **1996**, *382*, 559-561.
4. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature.* **1998**, *391*, 468-470.
5. Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. Structural basis for G•C recognition in the DNA minor groove. *Nat. Struct. Biol.* **1998**, *5*, 104-109.
6. Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science.* **1998**, *282*, 111-115.
7. Hsu, C. F., Phillips, J. W., Trauger, J. W., Farkas, M. E., Belitsky, J. M., Heckel, A., Olenyuk, B. Z., Puckett, J. W., Wang, C. C. C., and Dervan, P. B. Completion of a programmable DNA-binding small molecule library. *Tetrahedron.* **2007**, *63*, 6146-6151.
8. Belitsky, J. M., Leslie, S. J., Arora, P. S., Beerman, T. A., and Dervan, P. B. Cellular uptake of *N*-methylpyrrole/*N*-methylimidazole polyamide-dye conjugates. *Bioorg. Med. Chem.* **2002**, *10*, 3313-3318.
9. Crowley, K. S., Phillion, D. P., Woodard, S. S., Scheitzer, B. A., Singh, M., Shabany, H., Burnette, B., Hippenmeyer, P., Heitmeier, M., and Bashkin, J. K. Controlling the intracellular localization of fluorescent polyamide analogues in cultured cells. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1565-1570.
10. Best, T. P., Edelson, B. S., Nickols, N. G., and Dervan, P. B. Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12063-12068.
11. Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., and Dervan, P. B. Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic. Acids. Res.* **2004**, *32*, 2802-2818.
12. Xiao, X., Yu, P., Lim, H. S., Sikder, D., and Kodadek, T. A cell-permeable synthetic transcription factor mimic. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 2865-2868.
13. Nickols, N. G., Jacobs, C. S., Farkas, M. E., and Dervan, P. B. Improved nuclear localization of DNA-binding polyamides. *Nucleic. Acids. Res.* **2007**, *35*, 363-370.

14. Dose, C., Farkas, M. E., Chenoweth, D. M., and Dervan, P. B. Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit. *J. Am. Chem. Soc.* **2008**, *130*, 6859-6866.
15. Hsu, C. F., and Dervan, P. B. Quantitating the concentration of Py-Im polyamide-fluorescein conjugates in live cells. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5851-5855.
16. Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Molecular recognition of the nucleosomal "supergroove". *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6864-6869.
17. Dudouet, B., Burnett, R., Dickinson, L. A., Wood, M. R., Melander, C., Belitsky, J. M., Edelson, B., Wurtz, N., Briehn, C., Dervan, P. B., and Gottesfeld, J. M. Accessibility of nuclear chromatin by DNA binding polyamides. *Chem. Biol.* **2003**, *10*, 859-867.
18. Olenyuk, B. Z., Zhang, G. J., Klco, J. M., Nickols, N. G., Kaelin, Jr., W. G., and Dervan, P. B. Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16768-16773.
19. Kageyama, Y., Sugiyama, H., Ayame, H., Iwai, A., Fujii, Y., Huang, L. E., Kizaka-Kondoh, S., Hiraoka, M., and Kihara, K. Suppression of VEGF transcription in renal cell carcinoma cells by pyrrole-imidazole hairpin polyamides targeting the hypoxia responsive element. *Acta. Oncol.* **2006**, *45*, 317-324.
20. Nickols, N. G., Jacobs, C. S., Farkas, M. E., and Dervan, P. B. Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface. *ACS. Chem. Biol.* **2007**, *2*, 561-571.
21. Nickols, N. G., and Dervan, P. B. Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418-10423.
22. Matsuda, H., Fukuda, N., Ueno, T., Tahira, Y., Ayame, H., Zhang, W., Bando, T., Sugiyama, H., Saito, S., Matsumoto, K., and others, O. Development of gene silencing pyrrole-imidazole polyamide targeting the TGF-beta1 promoter for treatment of progressive renal diseases. *J. Am. Soc. Nephrology.* **2006**, *17*, 422-432.
23. Yao, E. H., Fukuda, N., Ueno, T., Matsuda, H., Matsumoto, K., Nagase, H., Matsumoto, Y., Takasaka, A., Serie, K., Sugiyama, H., and Sawamura, T. Novel gene silencer pyrrole-imidazole polyamide targeting lectin-like oxidized low-density lipoprotein receptor-1 attenuates restenosis of the artery after injury. *Hypertension.* **2008**, *52*, 86-92.
24. Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. Structural effects of DNA sequence on TA recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **2000**, *295*, 557-567.
25. Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* **2003**, *326*, 371-380.
26. Urbach, A. R., Love, J. J., Ross, S. A., and Dervan, P. B. Structure of a beta-alanine-linked polyamide bound to a full helical turn of purine tract DNA in the 1:1 motif. *J. Mol. Biol.* **2002**, *320*, 55-71.
27. Zhang, Q., Dwyer, T. J., Tsui, V., Case, D. A., Cho, J., Dervan, P. B., and Wemmer, D. E. NMR structure of a cyclic polyamide-DNA complex. *J. Am. Chem. Soc.* **2004**, *126*, 7958-7966.
28. Crothers, D. M., and Fried, M. Transmission of long-range effects in DNA. *Cold. Spring. Harb.*

*Symp. Quant. Biol.* **1983**, *47*, 263-269.

29. Lavelle, C. DNA torsional stress propagates through chromatin fiber and participates in transcriptional regulation. *Nat. Struct. Mol. Biol.* **2008**, *15*, 146-154.

30. Mathew-Fenn, R. S., Das, R., and Harbury, P. A. Remeasuring the double helix. *Science* **2008**, *322*, 446-449.

31. Nguyen-Hackley, D. H., Ramm, E., Taylor, C. M., Joung, J. K., Dervan, P. B., and Pabo, C. O. Allosteric inhibition of zinc-finger binding in the major groove of DNA by minor-groove binding ligands. *Biochemistry*. **2004**, *43*, 3880-3890.

32. Moretti, R., Donato, L. J., Brezinski, M. L., Stafford, R. L., Hoff, H., Thorson, J. S., Dervan, P. B., and Ansari, A. Z. Targeted chemical wedges reveal the role of allosteric DNA modulation in protein-DNA assembly. *ACS. Chem. Biol.* **2008**, *3*, 220-229.

33. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Acta. Crystallogr. D. Biol. Crystallogr.* **1992**, *26*.

34. Collaborative Computational Project, No. 4. The CCP4 suite: programs for protein crystallography. *Acta. Crystallogr.* **1994**, *50*, 760-763.

35. Schneider, T. R., and Sheldrick, G. M. Substructure solution with SHELXD. *Acta. Crystallogr. D. Biol. Crystallogr.* **2002**, *58*, 1772-1779.

36. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112-122 (2008).

37. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta. Crystallogr. D. Biol. Crystallogr.* **1997**, *53*, 240-255.

38. Emsley, P., and Cowtan, K. Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D. Biol. Crystallogr.* **2004**, *60*, 2126-2132.

39. Lavery, R., and Sklenar, H. Curves 5.2: Helical analysis of irregular nucleic acids. *Biochimie. Theorique, CNRS. URA.* **1997**, *77*.

40. Lu, X. J., Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids. Res.*, **31**, 5108-5121 (2003).

41. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.

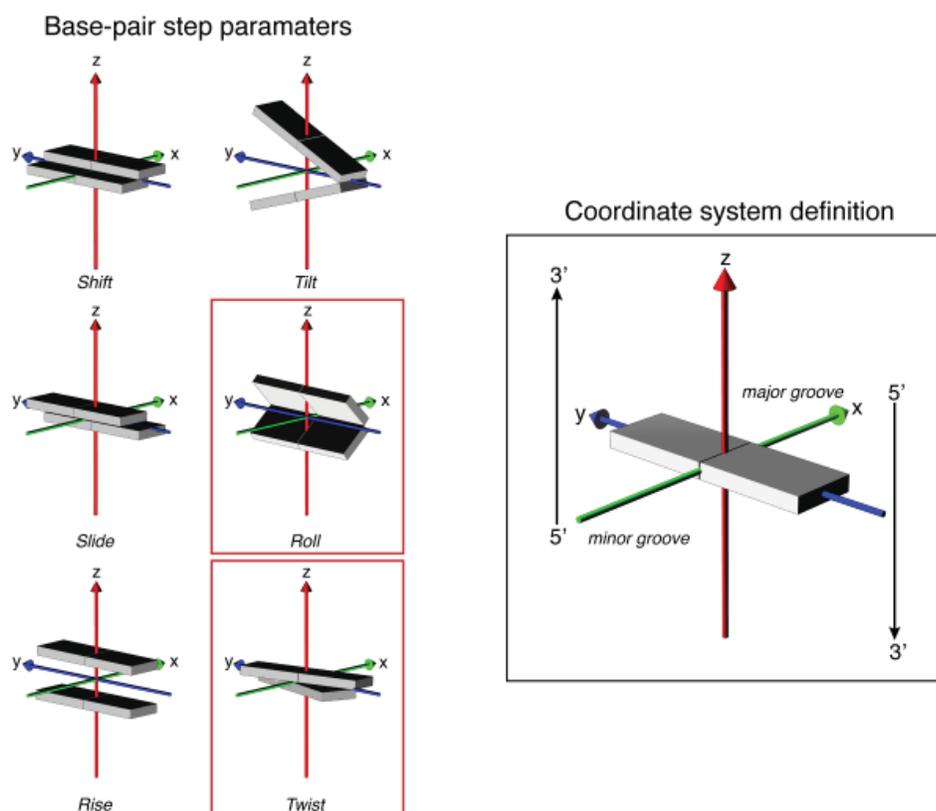
## 6.11 Supplemental Information

## Local base-pair step parameters

Parameter*	CC/GG	CA/TG	AG/CT	GT/AC	TA/TA	AC/GT	CT/AG	TG/CA	GG/CC
Shift, Å									
DNA	-0.58	0.23	0.58	-0.73	0.00	0.73	-0.58	-0.23	0.58
PA/DNA	-0.07	0.45	-1.72	-0.74	-0.04	0.78	1.75	-0.80	0.34
Slide, Å									
DNA	0.78	2.87	0.80	0.29	-0.10	0.29	0.80	2.87	0.78
PA/DNA	-0.12	0.86	0.26	-0.15	0.90	-0.21	0.28	1.29	0.13
Rise, Å									
DNA	3.28	3.29	3.10	3.29	3.66	3.29	3.10	3.29	3.28
PA/DNA	3.28	3.16	3.39	3.20	3.28	3.24	3.33	3.06	3.19
Tilt, °									
DNA	3.32	-1.37	-3.91	2.23	0.00	-2.22	3.90	1.37	-3.33
PA/DNA	5.81	0.74	-8.44	0.18	-0.16	0.40	7.94	0.64	-5.68
Roll, °									
DNA	9.03	-7.95	9.87	-0.74	3.26	-0.74	9.89	-7.96	9.03
PA/DNA	7.70	8.79	11.19	2.85	7.23	2.47	10.59	9.57	3.58
Twist, °									
DNA	28.26	50.50	21.04	35.80	48.48	35.80	21.04	50.49	28.26
PA/DNA	33.68	30.23	35.16	32.13	35.93	32.04	35.32	29.68	32.64

\*Relationship between the bases composing the base pair.

†DNA corresponds to 0.73 Å structure of duplex DNA solved by Rees and coworkers (PDB 1D8G, 5'-CCAGTACTGG-3').



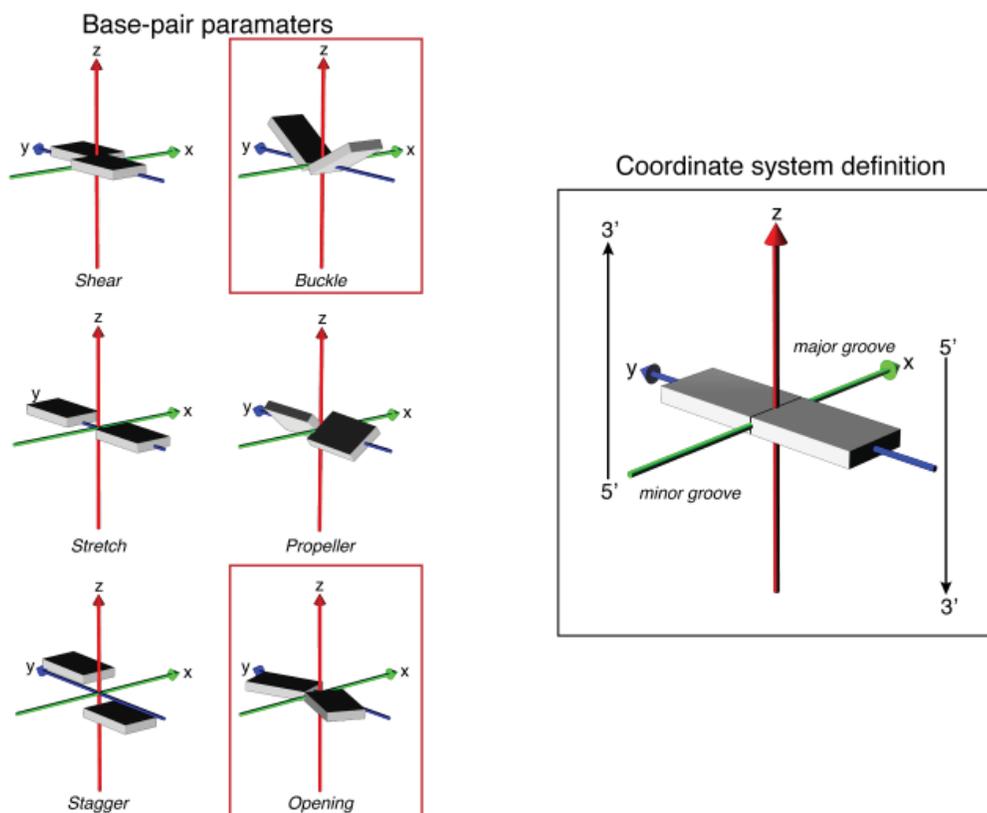
**Figure 6.6** Comparison of Local base-pair step parameters for DNA with and without polyamide complexed. Parameters were calculated using 3DNA.

## Local base-pair parameters and sugar pucker

Parameter*	C•G	C•G	A•T	G•C	T•A	A•T	C•G	T•A	G•C	G•C
Shear, Å										
DNA	0.15	0.32	0.02	-0.19	-0.09	0.09	0.19	-0.02	-0.32	-0.15
PA/DNA	0.15	0.19	0.13	-0.12	-0.13	0.13	0.12	-0.10	-0.16	-0.16
Stretch, Å										
DNA	-0.15	-0.16	-0.11	-0.07	-0.03	-0.03	-0.07	-0.11	-0.16	-0.15
PA/DNA	-0.14	-0.10	-0.11	-0.07	-0.10	-0.09	-0.10	-0.13	-0.04	-0.10
Stagger, Å										
DNA	0.09	0.12	0.08	0.25	0.00	0.00	0.25	0.08	0.12	0.09
PA/DNA	0.27	-0.02	-0.10	0.09	-0.02	0.04	0.09	-0.14	0.07	0.61
Buckle, °										
DNA	-6.41	-0.43	-1.63	10.75	13.15	-13.15	-10.74	1.63	0.44	6.42
PA/DNA	-10.59	-7.08	6.15	-2.23	0.37	1.17	2.39	-2.93	16.58	23.00
Propeller, °										
DNA	-15.16	-11.87	-3.23	-5.12	-10.90	-10.90	-5.15	-3.23	-11.87	-15.15
PA/DNA	-16.87	-6.31	-10.92	-4.05	-14.69	-16.26	-4.66	-7.96	-5.00	-17.05
Opening, °										
DNA	0.99	-1.15	1.35	1.26	-0.27	-0.27	1.27	1.34	-1.13	1.00
PA/DNA	0.68	1.71	3.84	-0.78	-4.87	-4.88	-1.10	3.78	0.87	-0.15
Sugar pucker										
DNA	O4'-endo C1'-exo	C2'-endo C3'-exo	C2'-endo C2'-endo	C1'-exo O4'-endo	C2'-endo C2'-endo	C2'-endo C2'-endo	O4'-endo C1'-exo	C2'-endo C2'-endo	C3'-exo C2'-endo	C1'-exo O4'-endo
PA/DNA	C3'-endo C2'-endo	C1'-exo C2'-endo	C2'-endo O4'-endo	C2'-endo C1'-exo	O4'-endo C1'-exo	C1'-exo O4'-endo	C1'-exo C2'-endo	O4'-endo C2'-endo	C2'-endo C2'-endo	C1'-exo C3'-endo

\*Relationship between the bases composing the base pair.

†DNA corresponds to 0.73 Å structure of duplex DNA solved by Rees and coworkers (PDB 1D8G, 5'-CCAGTACTGG-3').



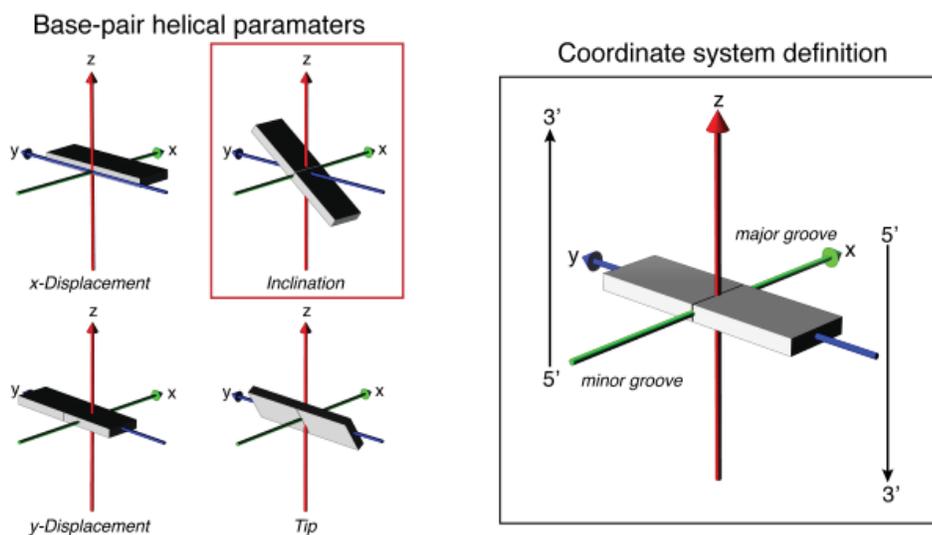
**Figure 6.7** Comparison of Local base-pair parameters and sugar conformations for DNA with and without polyamide complexed. Parameters were calculated using 3DNA.

**Local base-pair helical parameters**

Parameter*	CC/GG	CA/TG	AG/CT	GT/AC	TA/TA	AC/GT	CT/AG	TG/CA	GG/CC
X-displacement, Å									
DNA	-0.42	3.84	-1.31	0.57	-0.41	0.57	-1.32	3.84	-0.42
PA/DNA	-1.36	-0.07	-1.21	-0.76	0.42	-0.81	-1.07	0.59	-0.34
Y-displacement, Å									
DNA	1.83	-0.36	-2.71	1.50	0.00	-1.50	2.71	0.36	-1.84
PA/DNA	0.98	-0.68	1.45	1.36	0.05	-1.34	-1.59	1.60	-1.52
Inclination, °									
DNA	17.85	-9.25	25.08	-1.20	3.96	-1.20	25.12	-9.26	17.85
PA/DNA	12.96	16.43	17.68	5.14	11.58	4.47	16.74	18.10	6.29
Tip, °									
DNA	-6.56	1.59	9.92	-3.62	0.00	3.60	-9.90	-1.59	6.59
PA/DNA	-9.78	-1.38	13.33	-0.33	0.26	-0.73	-12.56	-1.20	9.97

\*Relationship between the bases composing the base pair.

†DNA corresponds to 0.73 Å structure of duplex DNA solved by Rees and coworkers (PDB 1D8G, 5'-CCAGTACTGG-3').



**Figure 6.8** Comparison of Local base-pair helical parameters for DNA with and without polyamide complexed. Parameters were calculated using 3DNA.

**Chapter 7: Programmable Oligomers Targeting 5'-GGGG-3' in the  
Minor Groove of DNA and NF- $\kappa$ B Binding Inhibition**

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

(Chenoweth, D.M., Puposki, J.A., Marques, M.A., Dervan, P. B. *Bioorg. Med. Chem.* **2007**, *15*, 759-770.)

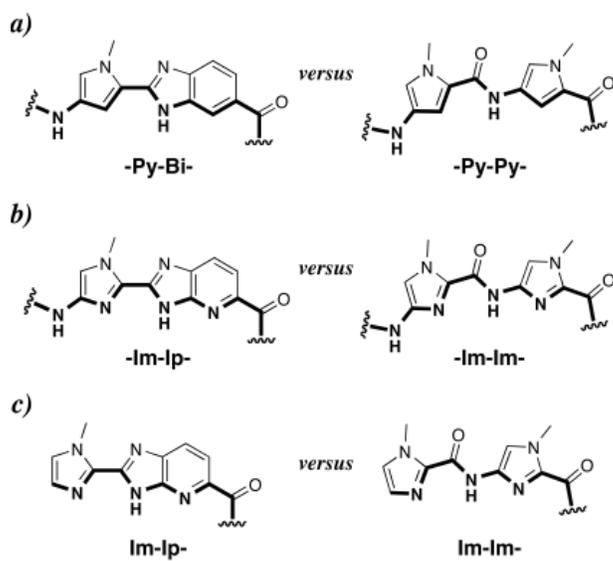
**Abstract**

A series of hairpin oligomers containing benzimidazole (Bi) and imidazopyridine (Ip) rings were synthesized and screened to target 5'-WGGGGW-3', a core sequence in the DNA-binding site of NF- $\kappa$ B, a prolific transcription factor important in biology and disease. Five Bi and Ip containing oligomers bound to the 5'-WGGGGW-3' site with high affinity. One of the oligomers (Im-Im-Im-Im- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp) was able to inhibit DNA binding by the transcription factor NF- $\kappa$ B.

## 7.1 Introduction

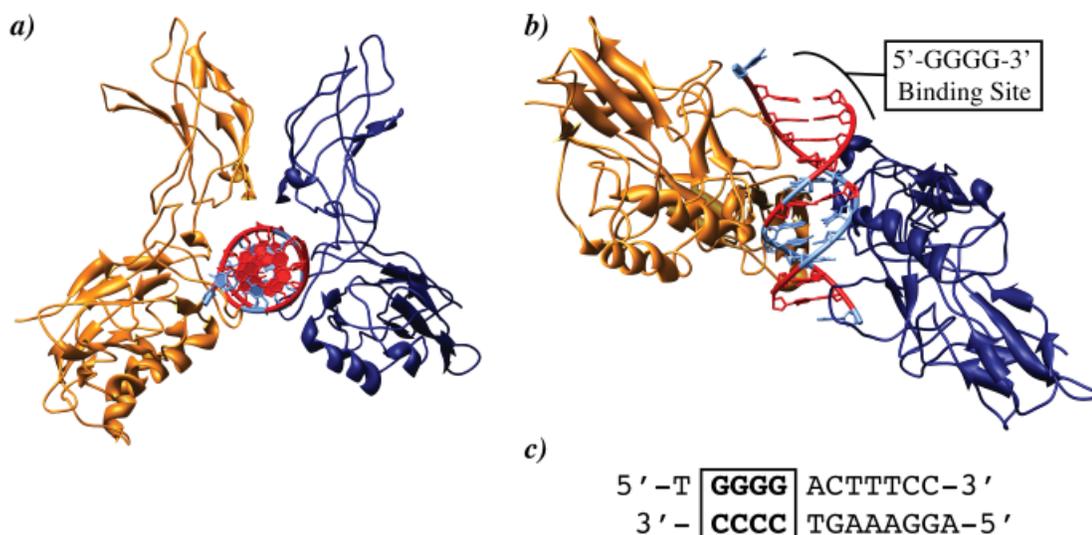
DNA-binding polyamides based on the architecture of the natural products netropsin and distamycin A have emerged as a promising class of gene modulators.<sup>1,2</sup> These molecules are capable of distinguishing the four Watson-Crick base pairs in the DNA minor groove and have been the subject of intense study along with other classes of minor groove binders for interfering with specific protein–DNA interfaces.<sup>3c,d-f,4,5</sup> Sequence-specific recognition of the minor groove of DNA by polyamides arises from the pairing of three different aromatic amino acids, pyrrole (Py), imidazole (Im), and hydroxypyrrole (Hp).<sup>4,5</sup> The direct readout, or information face, on the inside of the crescent-shaped polyamide may be programmed by the incremental change of atoms on the corners of the ring pairs presented to the DNA minor groove floor. Stabilizing and importantly, destabilizing interactions with the different edges of the four Watson-Crick bases are modulated by shape complementarity and specific hydrogen bonds.<sup>4-6,7</sup> For example, the Im/Py pair distinguishes G•C from C•G, T•A, and A•T. Im presents a lone pair of electrons to the DNA minor groove and can accept a hydrogen bond from the exocyclic amine of guanine.<sup>5</sup> Additionally, the Hp/Py pair distinguishes T•A from A•T, G•C, and C•G.<sup>4-6</sup> Hp projects an exocyclic OH group toward the minor groove floor that is sterically accommodated in the cleft of the T•A base pair, preferring to lie over T not A.<sup>5</sup> Since the development of these pairing rules based on the five-membered heterocyclic carboxamides Py, Im, and Hp, hundreds of DNA sequences have been successfully coded for using polyamides.<sup>3,5</sup> Yet, due to sequence dependent alterations in the shape of the DNA minor groove, there remain DNA sequences that prove difficult to target with high affinity and specificity.

There has been an ongoing effort to broaden the repertoire of heterocycles with improved DNA recognition, stability, and cellular trafficking profiles. We recently reported that the benzimidazole ring can be an effective platform for modular paired recognition elements in the minor groove of DNA.<sup>8,9</sup> The benzimidazole 6–5 bicyclic ring structure, while having slightly different curvature from the classic five-membered *N*-methyl pyrrole–carboxamides, presents an ‘inside edge’ with a similar atomic readout to the DNA minor groove floor, effectively mimicking Py, Im, and Hp (Figure 7.1).<sup>8,9</sup> The imidazopyridine/pyrrole pair Ip/Py distinguishes G•C from C•G and the hydroxybenzimidazole/pyrrole pair Hz/Py distinguishes T•A from A•T, providing a solution to the unanticipated hydroxypyrrole instability limitation.<sup>8,9</sup> This second generation solution to DNA recognition can be elaborated further, deleting incrementally almost all carboxamide linkages in the backbone of the hairpin motif creating an oligomer comprised of four dimer units capable of binding the site 5′-GTAC-3′, a sequence formally containing all four Watson-Crick base pairs.<sup>10</sup>



**Figure 7.1** Structures of the (a) pyrrole-benzimidazole internal dimer (-Py-Bi-), (b) imidazole-imidazopyridine internal dimer (-Im-Ip-), and (c) imidazole-imidazopyridine cap (Im-Ip-) in comparison with their respective five-membered ring systems. Hydrogen-bonding surfaces to the DNA minor-groove floor are bolded.

A key strategic issue for for small-molecule gene regulation is interfering with protein–DNA interfaces in the promoter of important genes. The new oligomer architecture has been successful in targeting the hypoxia inducible factor (HIF-1) binding site on the promoter of the VEGF gene.<sup>11</sup> This recent success spurred our interest in using these new oligomers to target another prolific transcription factor, NF- $\kappa$ B, important in biology and disease.<sup>12</sup> Guanine rich sequences are highly prevalent and partially conserved in the promoter region of NF- $\kappa$ B driven genes (Figure 7.2).<sup>12a,13</sup> Previous studies from our laboratory have established the ability of polyamides, targeting the sequence 5'-GGGACT-3', to inhibit DNA binding by NF- $\kappa$ B, however targeting the



**Figure 7.2** Crystal structure of the p50/p65 NF- $\kappa$ B heterodimer bound to the DNA duplex 5'-TG-GGGACTTTCC-3'.<sup>13</sup> The p50 and p65 monomers are represented as gold and dark blue ribbons, respectively. a) Top view looking down the DNA double helix. b) Side view showing the 5'-GGGG-3' oligomer binding site. GC rich regions are shown in red and AT rich regions are shown in light blue. c) Sequence of DNA bound to NF- $\kappa$ B.

four consecutive guanines was less successful due to the modest affinity of pyrrole/imidazole polyamides for this sequence.<sup>14,15</sup> The question arises whether a second generation oligomer architecture can target 5'-GGGG-3' recognition sequences with improved affinity.

We report the synthesis, DNA binding properties, and NF- $\kappa$ B:DNA binding inhibition properties of a new class of DNA binding oligomers targeted at the DNA sequence 5'-WGGGGW-3' containing 4-consecutive guanines (Figure 7.3). Oligomers vary by incorporation of benzimidazole-pyrrole into various positions in the parent polyamide **1**, resulting in oligomers **2-6**. From previous studies, we expect the BiPy dimer to be a good mimic for the traditional Py-Py recognition elements.<sup>16</sup> The Ip-Im moiety is introduced as a new mimic for the traditional Im-Im recognition motif (Figure 7.1). Quantitative DNase I footprinting titration experiments<sup>17</sup> were used to determine the binding affinities and specificities against single base-pair mismatch sites of oligomers **2-6** (Figure 7.3) as compared to their parent polyamide **1**. We found that in oligomers **2-6**, the 6-5 fused rings are effective mimics of their respective five-membered ring systems and that these oligomers target the binding-site 5'-WGGGGW-3' without loss of affinity as compared to parent hairpin polyamide **1**. In addition, we report the inhibition of DNA binding by the transcription factor NF- $\kappa$ B using this new class of minor groove binding oligomers (Figures 7.3 and 7.4).

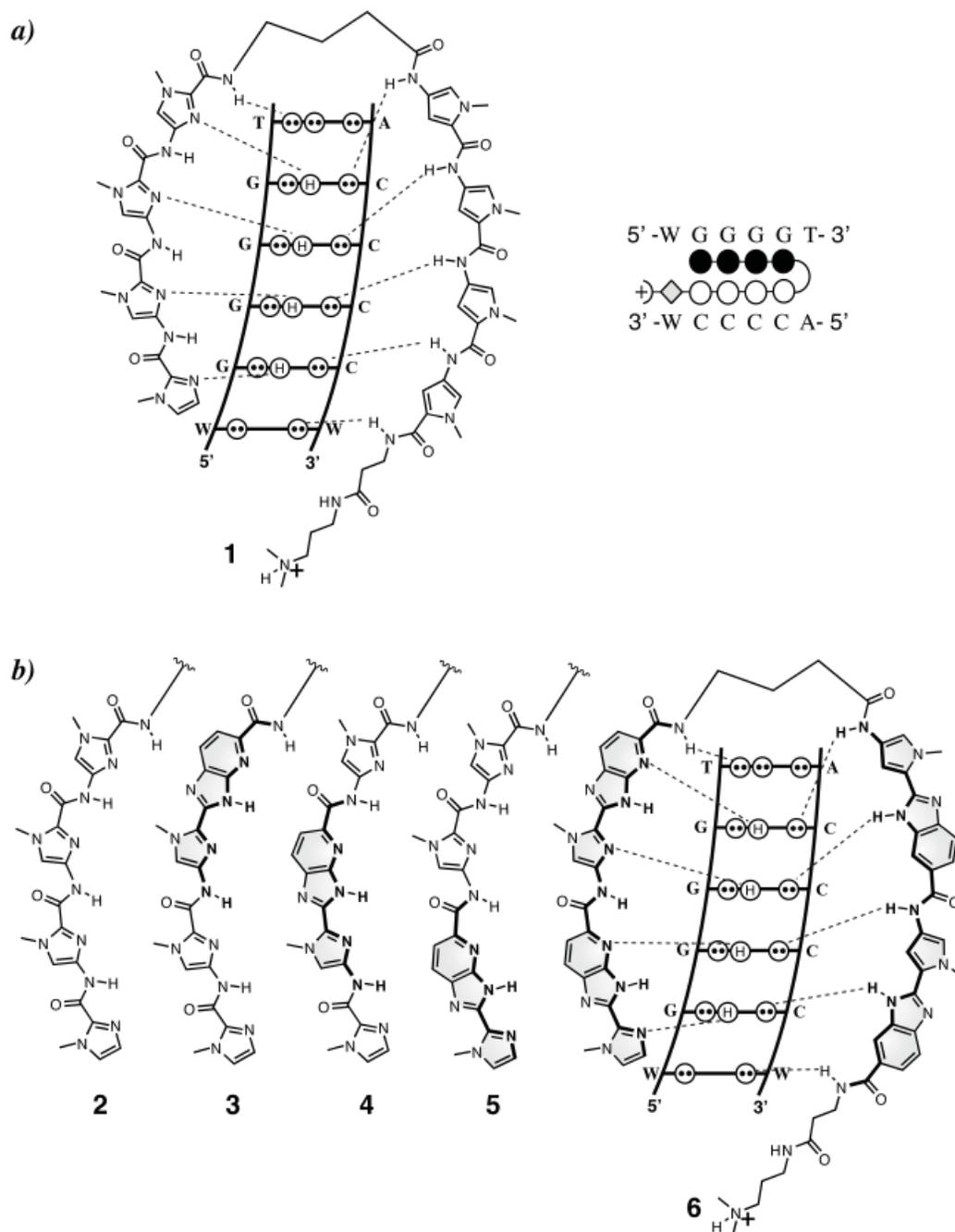
## 7.2 Results and Discussion

### 7.2.1 Heterocycle Synthesis

Dimeric units Im-Ip-OH (**9**) and Boc-Im-Ip-OH (**12**) were synthesized by oxidative condensation of aldehydes **7** and **10**, respectively, with previously reported diaminopyridine **8**. Mixing **7** and **8** in DMF at 80 °C for 1 h followed by 12 h of heating at 100 °C in the presence of FeCl<sub>3</sub> and air afforded **9** after purification by precipitation from water and saponification using a mixture of KOH (4 M) in MeOH at 40 °C (Scheme 7.1). Nitroimidazole (**10**) was synthesized from **7** using a mixture of H<sub>2</sub>SO<sub>4</sub> + SO<sub>3</sub> and neat red fuming nitric acid.<sup>18</sup> **10** was then added to a mixture of **8** in nitrobenzene and refluxed at 140 °C open to the atmosphere overnight to provide NO<sub>2</sub>-Im-Ip-OMe (**11**) after precipitation from water. Compound **11** was reduced using Pd/C in the presence of hydrogen followed by Boc protection using a mixture of Boc<sub>2</sub>O and DMAP in DMF to provide the final dimer Boc-Im-Ip-OH (**12**) after saponification and precipitation (Scheme 7.1).

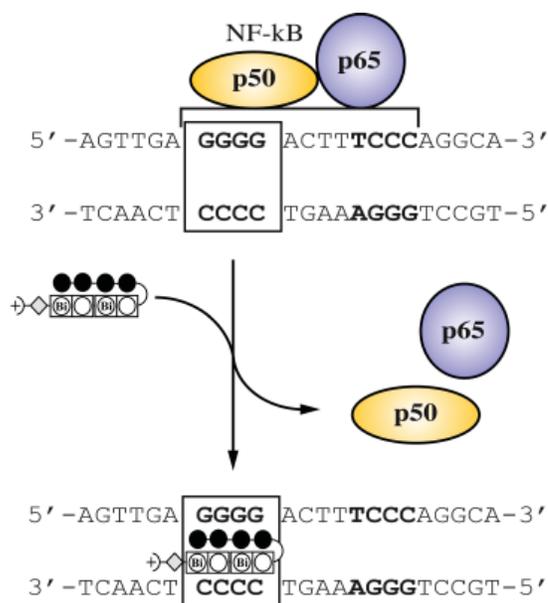
### 7.2.2 Oligomer Synthesis

Oligomers **1-6** were synthesized using manual solid phase synthesis methodology on



**Figure 7.3** Postulated hydrogen-bonding models for the 1:1 polyamide–DNA complexes with their matched sequence and the ball-and-stick representation for compounds **1** and **6** over the 6-base-pair matched binding site (variable region W = A or T). (a) Im-Im-Im-Im- $\gamma$ -Py-Py-Py-Py- $\beta$ -Dp (**1**), (b) Im-Im-Im-Im- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp (**2**), Im-Im-Im-Ip- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp (**3**), Im-Im-Ip-Im- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp (**4**), Im-Ip-Im-Im- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp (**5**), Im-Ip-Im-Ip- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp (**6**).

commercially available  $\beta$ -Ala-PAM resin as previously described (Scheme 7.2).<sup>19</sup> Starting from base resin (**BR1**), monomeric and dimeric heterocyclic units were appended onto the resin in stepwise fashion using HBTU activation. Couplings were allowed to proceed for several h between

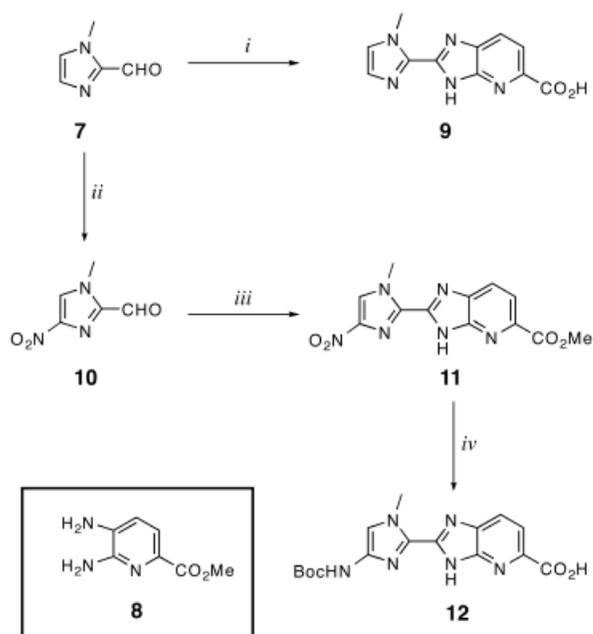


**Figure 7.4** Illustration of NF-κB:DNA binding inhibition by oligomer 2.

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

### 7.2.3 DNA affinity and sequence specificity

Quantitative DNase I footprinting titrations were carried out for oligomers **1–6** on the PCR product of plasmid pEF16 (Figures 7.5, 7.6, and 7.7). Plasmid pEF16 was constructed containing two designed match sites (5'-XGGGGT-3' X = A, T) and two mismatch sites (5'-AGGGAT-3' and 5'-AGAGGT-3') (Figures 7.5, 7.6, and 7.7). The first two match sites 5'-TGGGGT-3' and 5'-AGGGGT-3' were included to determine if there was an energetic penalty associated with a 5'-AG-3' step. Previously, DNA sequences containing multiple 5'-AG-3' transitions have proven more difficult to target at high affinity and it has been shown that changes in flanking sequence can affect binding at a proximal site.<sup>20</sup> The second two binding sites, 5'-AGGGAT-3' and 5'-AGAGGT-3', were designed to elucidate the energetic penalty for the loss of a favorable hydrogen bond between the exocyclic amine of guanine and the lone pair nitrogen on the oligomer in question. Control compound **1** bound both 4-G match sequences (5'-XGGGGT-3', X = A, T) with comparably low affinity ( $K_a \sim 10^8 \text{ M}^{-1}$ ), showing no bias for either site (Table 7.1). Compound **1** distinguished against mismatch sequences (5'-AGGGAT-3' and 5'-AGAGGT-3') with roughly 10-fold specificity (Table 7.1). Oligomer **2** demonstrated a large increase in affinity ( $K_a \sim 10^{10} \text{ M}^{-1}$ ) as compared to **1** for both match sites, but showed lower specificity (4-fold) over the mismatch



**Scheme 7.1.** Synthesis of imidazopyridine–imidazole dimers Im-*Ip*-OH (**9**) and Boc-Im-*Ip*-OH (**12**). Reagents and conditions: (i) **8**, PhNO<sub>2</sub>, 140 °C; (ii) oleum, red-fuming HNO<sub>3</sub>; (iii) **8**, PhNO<sub>2</sub>, 140 °C; (iv) H<sub>2</sub>, Pd/C, DMF; (Boc)<sub>2</sub>O, DIEA, DMAP, DMF, 4 N KOH, MeOH.

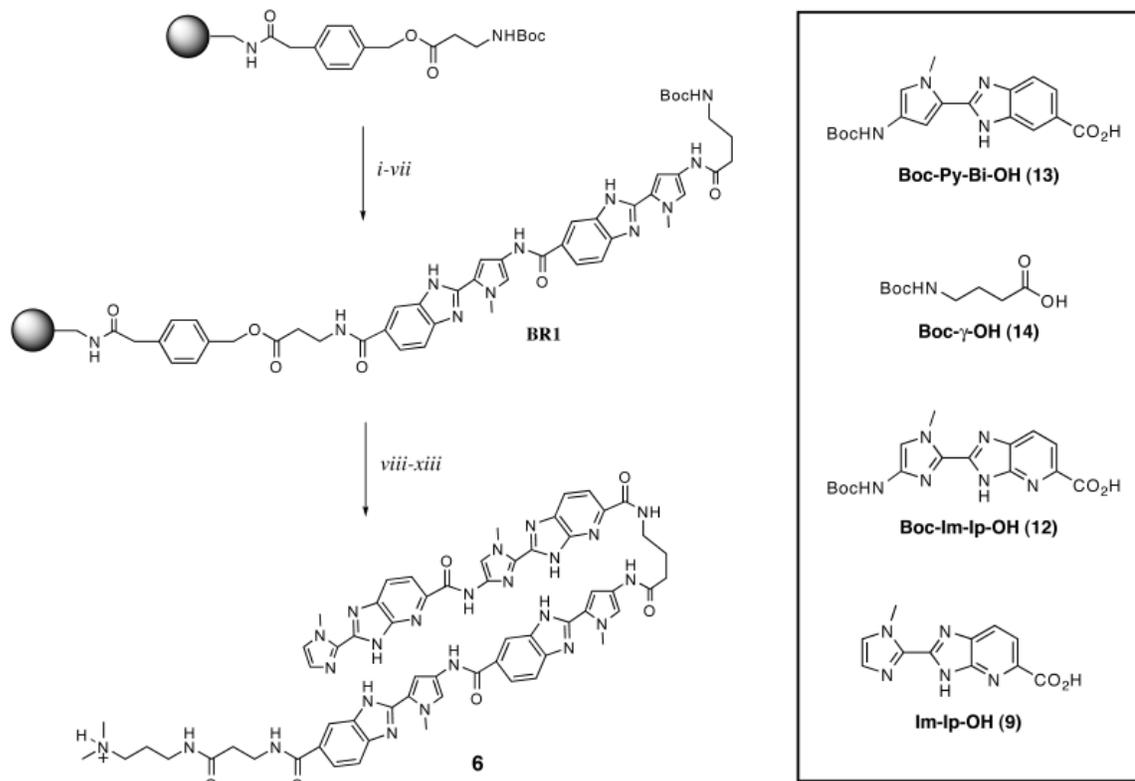
placement in the DNA minor groove.<sup>21</sup> Aromatic stacking and van der Waals interactions are also contributing factors.<sup>22,23</sup> Furthermore, the benzimidazole derivatives are a more rigid structure with a lower degree of rotational freedom. Such pre-organization may decrease the entropic cost of DNA complexation.

#### 7.2.4 NF- $\kappa$ B electrophoretic mobility gel shift assay

The guanine rich region of the NF- $\kappa$ B binding element has been shown to be important in protein–DNA recognition (Figure 7.2).<sup>13,15,24,25</sup> Polyamides targeted to this region have previously been shown to inhibit NF- $\kappa$ B binding.<sup>15</sup> The possibility of steric or allosteric inhibition exists due to numerous contacts between the p50 protein and the phosphate backbone or direct protein–DNA contacts in the major groove, opposite the minor groove polyamide binding site.<sup>13,15</sup> To test whether the second generation oligomer architecture would be successful at targeting 5'-GGGG-3' in a biologically relevant context we employed an NF- $\kappa$ B electrophoretic mobility gel shift assay. The

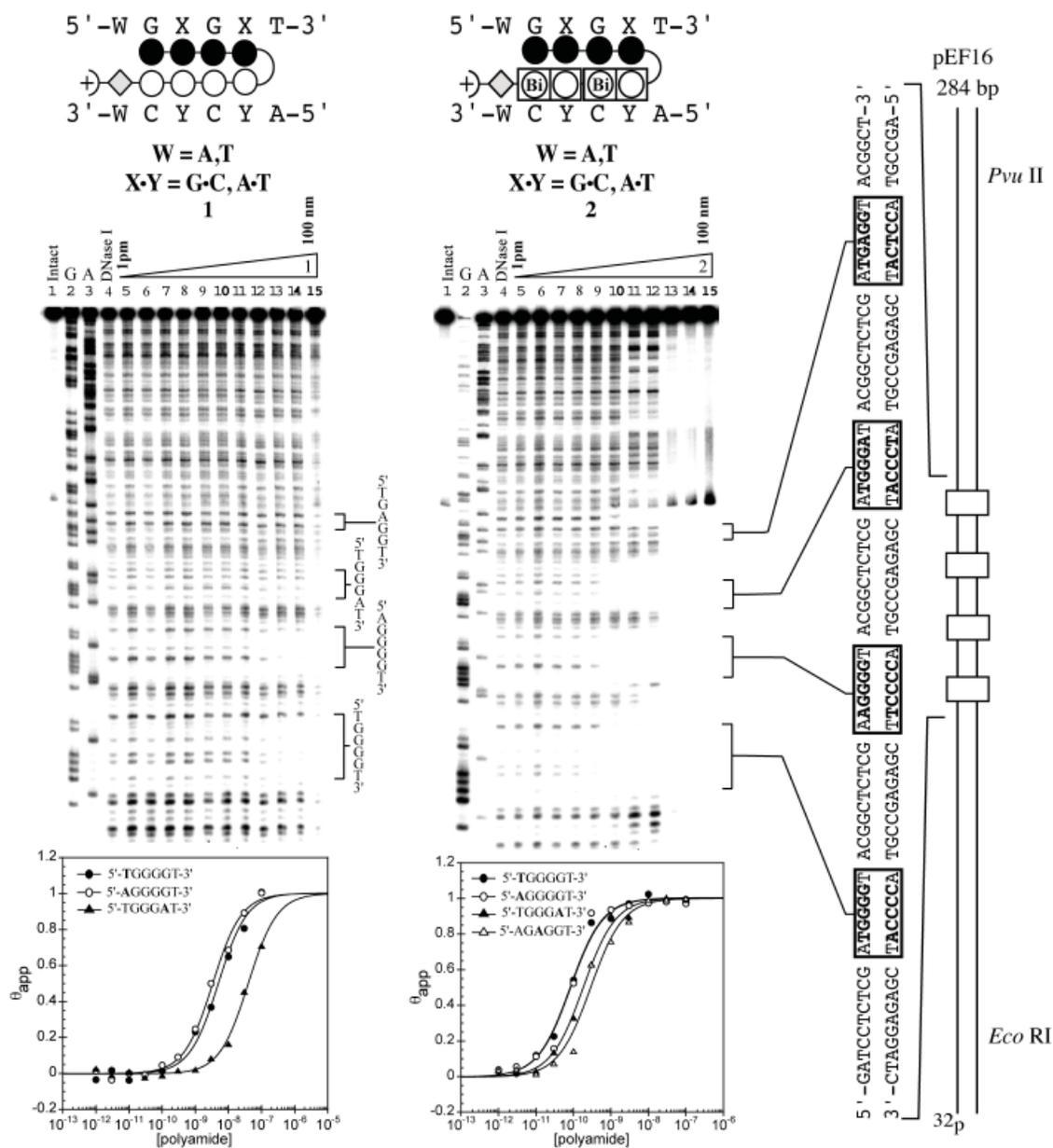
sites. Oligomer **5** showed a moderate increase in affinity ( $K_a \sim 10^9$  M<sup>-1</sup>) and demonstrated minor selectivity (3-fold) over the mismatch 5'-atGGGAt-3', however showed an 11-fold specificity over the mismatch 5'-atGAGGt-3'. Oligomer **4** demonstrated high affinity for the 4-G match sequences ( $K_a \sim$  mid  $10^9$  M<sup>-1</sup>) with reasonable 5-fold selectivity over the mismatch sequences (5'-AGGGAT-3' and 5'-AGGAGT-3'). Compounds **3** and **6** bound all designed sequences with similar affinity. Thermodynamic data for oligomers **1–6** are summarized in Table 7.1.

In general, the global increase in affinity for these novel oligomers is not altogether unexpected. In contrast to the 5-membered heterocyclic carboxamides, the 6–5 fused benzimidazole analogues have a larger hydrophobic surface, likely promoting their

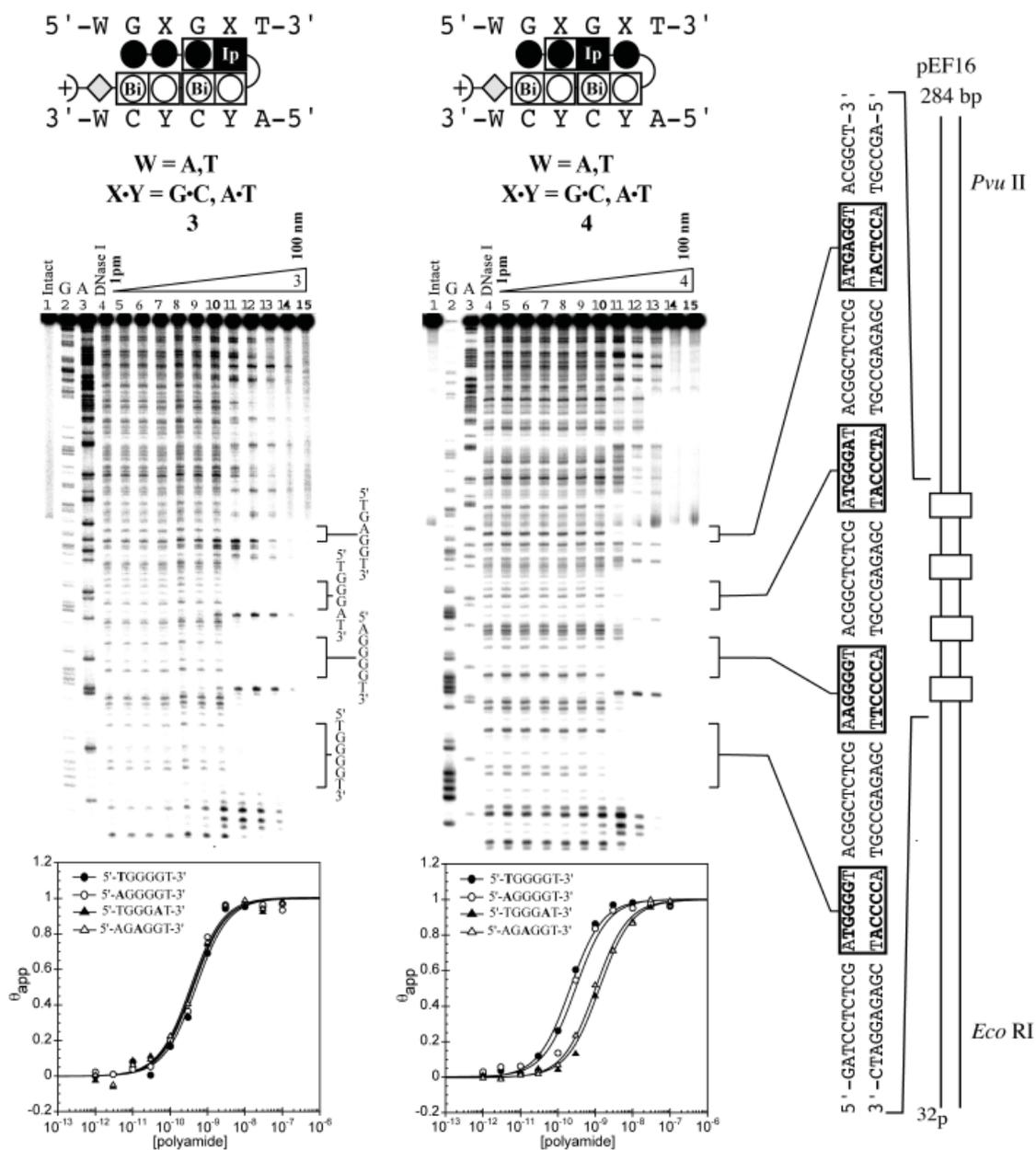


**Scheme 7.2.** Representative solid-phase synthesis of polyamide **6** along with a table of the amino acid building blocks used for the synthesis. Reagents and conditions: (i) 80% TFA/DCM; (ii) Boc-Py-Bi-OH (**13**), HBTU, DIEA, DMF; (iii) Ac<sub>2</sub>O, DIEA, DMF; (iv) repeat (i–iii); (v) 80% TFA/DCM; (vi) Boc-c-OH (**14**), HBTU, DIEA, DMF; (vii) Ac<sub>2</sub>O, DIEA, DMF to provide **BR1**; (viii) 80% TFA/DCM; (ix) Boc-Im-Ip-OH (**12**), HBTU, DIEA, DMF; (x) 80% TFA/DCM; (xi) Im-Ip-OH, HBTU, DIEA, DMF; (xii) dimethylaminopropylamine (Dp), 80 °C 2 h; (xiii) preparative HPLC to give **6**.

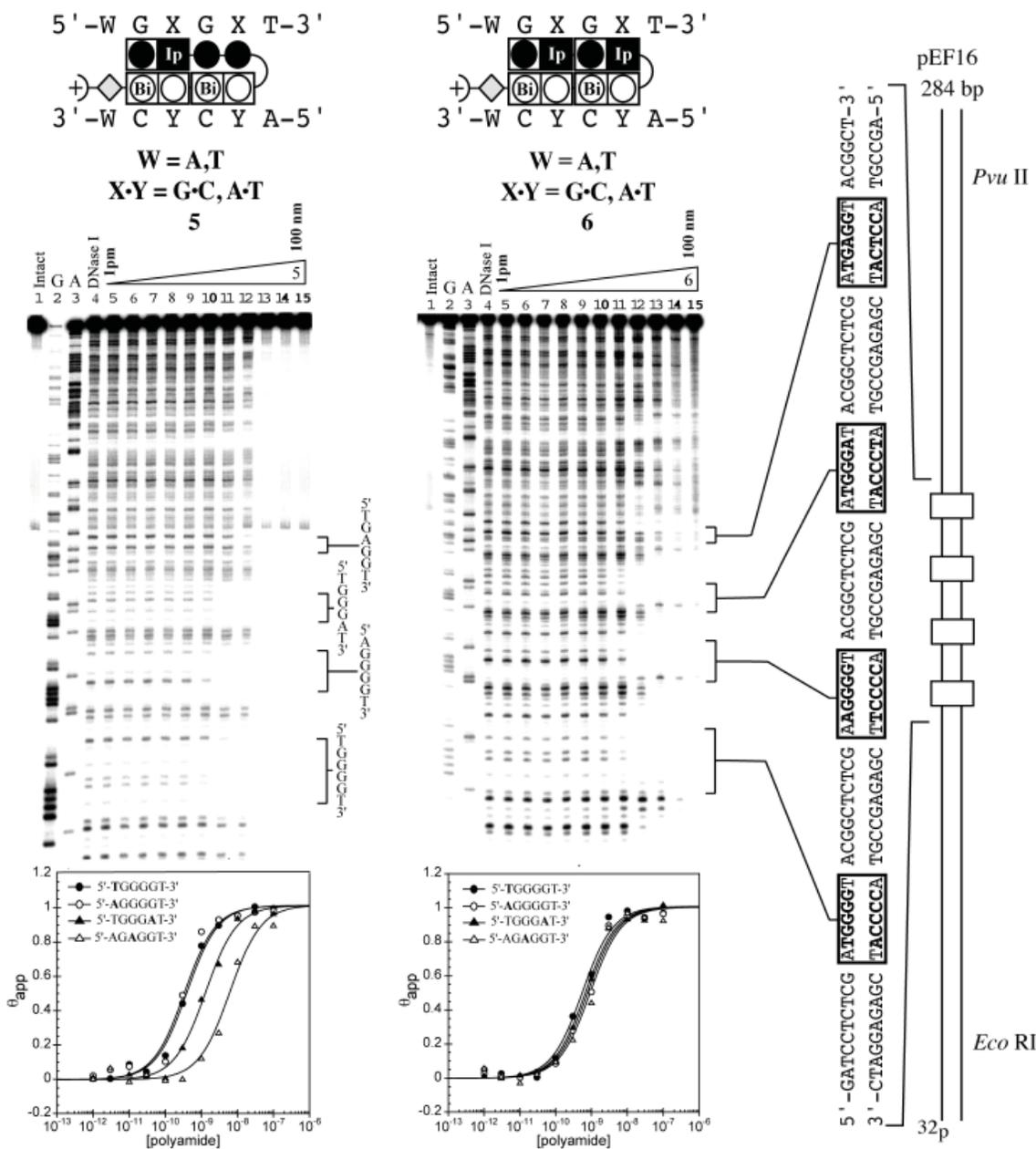
NF-κB binding inhibition properties of compounds **1–6** were screened against a 40 bp DNA probe containing the 5'-WGGGGW-3' sequence, which is part of an element from the intronic enhancer of the immunoglobulin κ light-chain gene recognized and bound by NF-κB (Figure 7.8).<sup>13</sup> In the initial screen each oligomer was tested at two concentrations, 10 and 100 nM, for the ability to interfere with protein binding (Figure 7.8). At the higher concentration, **2** and **4** demonstrate a clear decrease in band intensity, whereas compounds **1**, **3**, **5**, and **6** have little effect. This result agrees with the footprinting data, as protein inhibition appears to scale with compound affinity for the 5'-WGGGGW-3' site. At the 10 nM concentration compound **2** is the only compound to show significant protein inhibition, reducing band intensity by more than 80% (Figure 7.8). Full titration of compound **2** over a concentration range of 500 pM–500 nM established an EC<sub>50</sub> of 15.7 nM for NF-κB inhibition as shown in Figure 7.9. In addition, the identity of the NF-κB band was confirmed by antibody supershift and the data are shown in Figure 7.10.



**Figure 7.5** Quantitative DNaseI footprinting experiments in the hairpin motif for polyamides **1** and **2** respectively, on the 278 bp, 5'-end-labeled PCR product of plasmid pEF16: lane 1, intact DNA; lanes 2 and 3, G and A sequencing reaction; lane 4, DNase I standard; lanes 5–15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM polyamide concentration, respectively. Each footprinting gel is accompanied by the following: (top) ball-and-stick models of the compound bound to its target DNA sequence; and (bottom) Binding isotherms for the four designed sites.  $\theta_{app}$  values were obtained according to published methods. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; dimer units are represented as rectangles containing either Bi enclosed in a circle representing benzimidazole or Ip in a black box representing imidazopyridine; beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

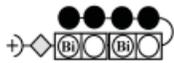


**Figure 7.6** Quantitative DNaseI footprinting experiments in the hairpin motif for polyamides **3** and **4** respectively, on the 278 bp, 5'-end-labeled PCR product of plasmid pEF16: lane 1, intact DNA; lanes 2 and 3, G and A sequencing reaction; lane 4, DNase I standard; lanes 5–15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM polyamide concentration, respectively. Each footprinting gel is accompanied by the following: (top) ball-and-stick models of the compound bound to its target DNA sequence; and (bottom) Binding isotherms for the four designed sites.  $\theta_{app}$  values were obtained according to published methods. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; dimer units are represented as rectangles containing either Bi enclosed in a circle representing benzimidazole or Ip in a black box representing imidazopyridine; beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.



**Figure 7.7** Quantitative DNaseI footprinting experiments in the hairpin motif for polyamides **5** and **6** respectively, on the 278 bp, 5'-end-labeled PCR product of plasmid pEF16: lane 1, intact DNA; lanes 2 and 3, G and A sequencing reaction; lane 4, DNase I standard; lanes 5–15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM polyamide concentration, respectively. Each footprinting gel is accompanied by the following: (top) ball-and-stick models of the compound bound to its target DNA sequence; and (bottom) Binding isotherms for the four designed sites.  $\theta_{app}$  values were obtained according to published methods. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; dimer units are represented as rectangles containing either Bi enclosed in a circle representing benzimidazole or Ip in a black box representing imidazopyridine; beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

**Table 7.1** Affinities of 5'-GGGG-3' binding oligomers  $K_a$  ( $M^{-1}$ )<sup>a,b</sup>

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

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses.

b) Assays were performed at 22°C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0.

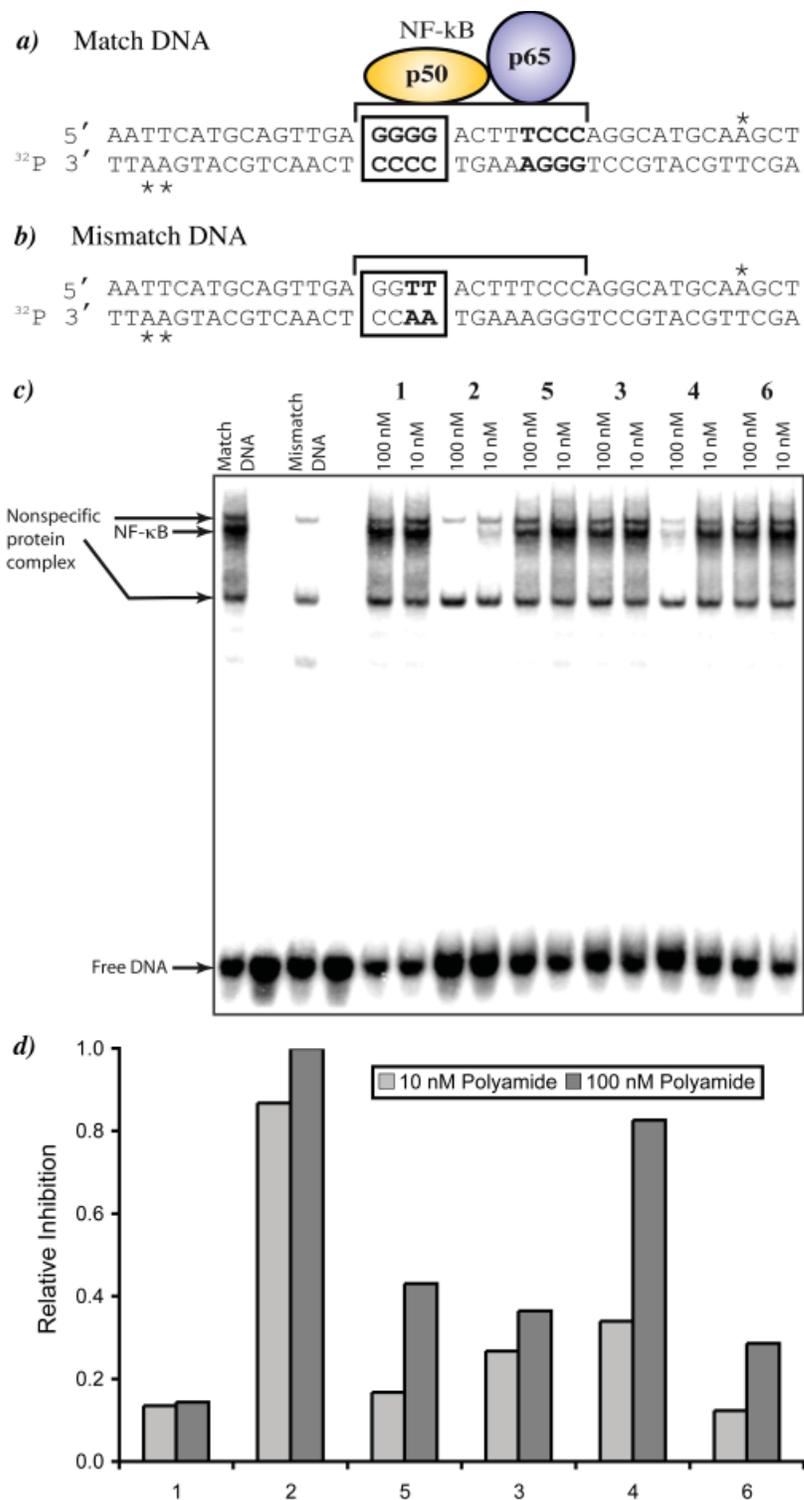
### 7.3 Conclusion

While traditional Py, Im, and Hp containing polyamides have been successful at recognizing hundreds of pre-determined DNA sequences with high affinity and specificity, a host of target sequences such as 5'-GGGG-3' have proven difficult to code for using polyamides. A series of novel oligomers containing the 6–5 fused benzimidazole (Bi) and imidazopyridine (Ip) heterocycles were developed. These oligomers, composed of 5-membered heterocyclic carboxamides, demonstrated a substantial increase in affinity (10- to 100-fold) for the 5'-GGGG-3' sequence. The marked increase in affinity could be attributed to a combination of oligomer properties including a larger hydrophobic surface, a high degree of ligand pre-organization, or differential solvation/desolvation effects. The ability of this class of new oligomers to inhibit protein–DNA binding was demonstrated by the inhibition of NF-κB. We are encouraged by the fact that these oligomers demonstrate improved affinity for guanine rich DNA sequences and future work directed toward improving sequence specificity and examination of the nuclear trafficking ability is a priority.

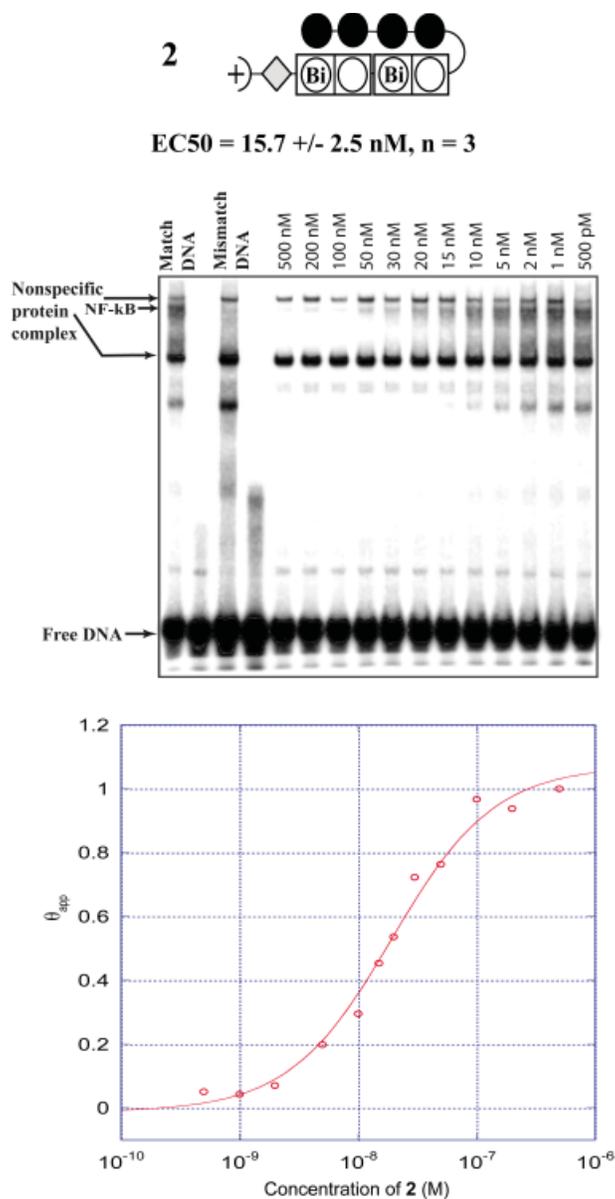
### 7.4 Experimental

#### 7.4.1 General

*N,N*-Dimethylformamide (DMF), *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylaminopropylamine (Dp), triethylamine (TEA), nitrobenzene (NO<sub>2</sub>Ph), 2-formyl-N-

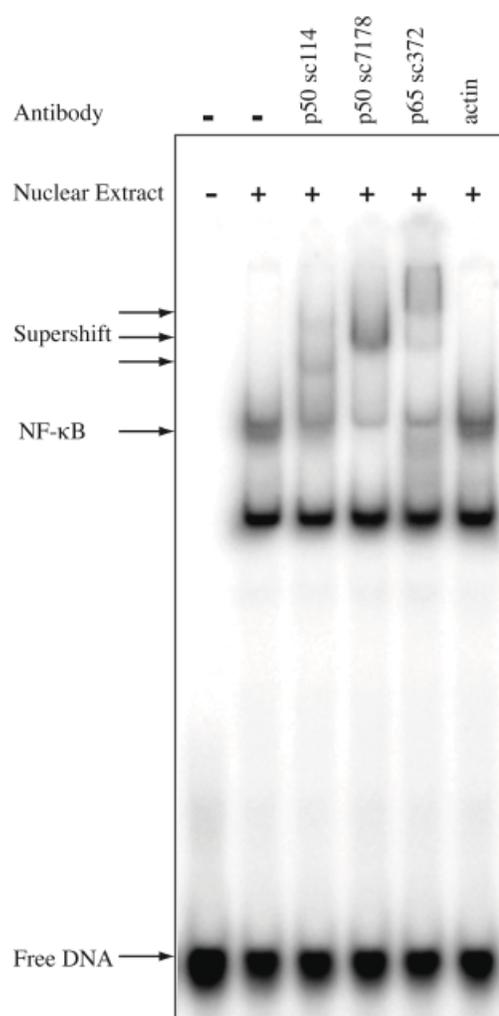


**Figure 7.8** a) Match DNA sequence with the p50 protein of NF-κB overlapping the oligomer binding site. Asterisks indicate the location of radio-labeled nucleotides in the probe sequence. b) Mismatch DNA sequence. c) Gel shift screen for compounds **1–6** at concentrations of 10 and 100 nM. d) Plot of relative NF-κB inhibition for compounds **1–6**.



**Figure 7.9** (Top) ball-and-stick model for **2** and  $EC_{50}$  value from gel shift experiment. (Middle) representative NF- $\kappa$ B titration gel ( $n = 3$ ) for **2**. (Bottom) binding isotherm for **2**.

methylimidazole, red fuming nitric acid, 1,3-dichloro-4-nitropyridine, 30% bromine in acetic acid, palladium acetate  $Pd(OAc)_2$ , and 10% palladium on carbon were purchased from Aldrich. Boc- $\beta$ -alanine-(4-carboxylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- $\beta$ -Pam-resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), *N,N*-dimethylaminopyridine (DMAP), and Boc- $\beta$ -alanine were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was



**Figure 7.10** Antibody supershift on match DNA. The NF- $\kappa$ B band is shifted in the presence of anti-p50 or anti-p65 antibody.

purchased from Halocarbon. All other solvents were reagent grade from EM. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Precoated silica gel plates 60F<sub>254</sub> for TLC and silica gel 60 (40 μm) for flash chromatography were from Merck. Glycogen (20 mg/ml), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer–Mannheim. pUC19 was purchased from New England Biolabs and deoxyadenosine [ $\gamma$ -<sup>32</sup>P] triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) was obtained from Amersham Pharmacia. DNase I (7500 U/ml, FPLC pure) was purchased from Roche. AmpliTaq DNA polymerase was from Perkin-Elmer and used with the provided buffers. Tris–HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris–borate–EDTA was from Gibco and bromophenol blue was from Acros. All reagents were used without further purification. NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-d<sub>6</sub> or CDCl<sub>3</sub> with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett–Packard (model 8452 A) diode array spectrophotometer. High resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the California Institute of Technology. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) was conducted at the Mass Spectroscopy Laboratory at the California Institute of Technology.

#### 7.4.2 Heterocycle Synthesis

Heterocyclic building blocks Boc-Im-OH and Boc-Py-Bi-OH were synthesized as reported.<sup>8,19</sup> Im-Im-OH (1-methyl-4-(1-methyl-1*H*-imidazole-2-carboxamido)-1*H*-imidazole-2-carboxylic acid, CAS [464892-44-2]) and Boc-Im-Im-OH (4-(4-(tert-butoxycarbonylamino)-1-methyl-1*H*-imidazole-2-carboxamido)-1-methyl-1*H*-imidazole-2-carboxylic acid, CAS [502170-49-2]), are commercially available.

##### 7.4.2.1 1-methyl-4-nitro-1*H*-imidazole-2-carbaldehyde (NO<sub>2</sub>-Im-CHO) (**10**).

A cooled flask (0 °C) of 1-methyl-2-imidazole-carboxaldehyde (**7**) (8g, 72.6 mmol, Aldrich) was treated dropwise with a precooled (0 °C) solution of red fuming nitric acid (75 ml) in conc. H<sub>2</sub>SO<sub>4</sub>•SO<sub>3</sub> (30%) (75 ml). The mixture was warmed to room temperature and stirred for 12 h open to the atmosphere. Next, the mixture was poured over ice, neutralized with solid sodium carbonate, extracted 4 times with dichloromethane, dried over anhydrous sodium sulfate, and concentrated in

vacuo to give a brownish-yellow oil. The oil was recrystallized from *i*PrOH/Et<sub>2</sub>O or EtOH/Et<sub>2</sub>O to give 1-methyl-4-nitro-1*H*-imidazole-2-carbaldehyde (**10**) as a tan crystalline solid (4.5 g, 40% Yield). TLC (1:1 EtOAc/Hex) *R<sub>f</sub>* = 0.4; <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>) δ 9.74 (s, 1H), 8.71 (s, 1H), 3.99 (s, 3H); <sup>13</sup>C (75 MHz, DMSO-*d*<sub>6</sub>) δ 182.31, 146.14, 140.56, 127.33, 35.70; HR-MS (EI<sup>+</sup>): calculated for C<sub>5</sub>H<sub>5</sub>O<sub>3</sub>N<sub>3</sub>: 155.0330; found: 155.0350.

7.4.2.2 2-(1-Methyl-4-nitroimidazol-2-yl)-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester (NO<sub>2</sub>-*Im-Ip-OMe*) (**11**).

1-Methyl-4-nitro-1*H*-imidazole-2-carbaldehyde (2.01 g, 13.0 mmol) (**10**) and methyl 5,6-diaminopyridine-2-carboxylate (**8**) (2.17 g, 13.0 mmol) suspended in 120 ml of nitrobenzene was heated to 140 °C for 48 h open to the atmosphere. The reaction mixture was cooled to 23 °C and the precipitate collected by vacuum filtration. The solid was washed with diethyl ether and dried under high vacuum to provide 2-(1-methyl-4-nitroimidazol-2-yl)-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester (**11**) (3.6 g, 92% Yield) as a powdery tan solid. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>) δ 14.02 (broad s, 1H), 8.03-8.06 (m, 2H), 4.27 (s, 3H), 3.91 (s, 3H); <sup>13</sup>C (75 MHz, DMSO-*d*<sub>6</sub>) δ 166.01, 146.44, 142.95, 136.59, 135.80, 130.36, 127.35, 123.81, 121.17, 121.04, 52.85, 37.43; HR-MS (EI<sup>+</sup>): calculated for C<sub>12</sub>H<sub>10</sub>N<sub>6</sub>O<sub>4</sub>: 302.0760; found: 302.0760.

7.4.2.3 2-{4-Amino-1-methylimidazol-2-yl}-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester (H<sub>2</sub>N-*Im-Ip-OMe*).

2-(1-Methyl-4-nitroimidazol-2-yl)-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester (**11**) (3 g, 9.9 mmol) was dissolved in anhydrous DMF (150 ml) and the solution was degassed with Ar. After the addition of Pd/C (10 wt. %, 600 mg) the reaction mixture was purged 3 times with hydrogen and then left to stir at 23°C for 9 h under a hydrogen balloon atmosphere. After filtering through a pad of Celite and washing with copious amounts EtOAc the filtrate was concentrated in vacuo to give 2-{4-amino-1-methylimidazol-2-yl}-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester, without further purification (2.7 g, 100% Yield). <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>) δ 166.17, 149.03, 148.21, 141.40, 131.50, 120.12, 107.77; HR-MS (EI<sup>+</sup>): calculated for C<sub>12</sub>H<sub>12</sub>N<sub>6</sub>O<sub>2</sub>: 272.1020; found: 272.1030.

7.4.2.4 2-{4-[(*tert*-Butoxy)carbonylamino]-1-methylimidazol-2-yl}-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester (*Boc-Im-Ip-OMe*).

2-{4-amino-1-methylimidazol-2-yl)-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester (**11**) (2.0 g, 7.3 mmol) dissolved in DMF (25 ml) was treated with Boc<sub>2</sub>O (5.3 g, 24.3 mmol), DIEA (5.2 ml), and DMAP (95 mg, 0.73 mmol). The reaction mixture was then heated to 80 °C for 72 h, cooled to 23 °C, and flashed through a plug of silica gel eluting with EtOAc to give a mixture of mono- and di-boced (2-{4-[(*tert*-butoxy)carbonylamino]-1-methylimidazol-2-yl)-3-[(*tert*-butoxy)carbonylamino]-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester) products which were carried on for saponification.

*7.4.2.5 2-{4-[(*tert*-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic (Boc-Im-Ip-OH) (12).*

2-{4-[(*tert*-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3-[(*tert*-butoxy)carbonylamino]-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester dissolved in MeOH (10 ml) and NaOH (1 N, 25 ml) was heated to 50 °C for 4 h. The reaction mixture was cooled to 0 °C and the pH adjusted slowly to pH = 4 with 1 N HCl. The reaction mixture was then extracted with ethyl acetate (4 times), dried over anhydrous sodium sulfate, concentrated in vacuo, and dried under high vacuum to give 2-{4-[(*tert*-butoxy)carbonylamino]-1-methylimidazol-2-yl)-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic (**12**) (258 mg, 60% Yield) as a light yellow solid. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>) δ 13.19 (broad s, 2H), 9.53 (s, 1H), 8.00 (m, 2H), 7.36 (s, 1H), 4.15 (s, 3H), 1.47 (s, 9H); <sup>13</sup>C (75 MHz, DMSO-*d*<sub>6</sub>) δ 166.61, 152.92, 147.81, 142.48, 138.38, 132.01, 128.91, 119.74, 113.50, 79.07, 35.35, 28.12; HR-MS (EI<sup>+</sup>): calculated for C<sub>16</sub>H<sub>18</sub>N<sub>6</sub>O<sub>4</sub>: 358.1390; found: 358.1370.

*7.4.2.6 2-(1-Methylimidazol-2-yl)-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester (Im-Ip-OMe).*

1-Methylimidazole-2-carbaldehyde (**7**) (214 mg, 1.85 mmol) and methyl 5,6-diaminopyridine-2-carboxylate (**8**) (310 mg, 1.85 mmol) were suspended in 17 ml of DMF and heated to 80 °C for 1 hour open to the atmosphere. Next, FeCl<sub>3</sub>•6H<sub>2</sub>O (24mg, 0.09 mmol) was added and the reaction mixture was heated to 100 °C for 12 hrs while air was bubbled through the reaction mixture. The reaction mixture was then cooled to room temperature and poured over ice. The precipitate was collected by filtration and washed with cold diethyl ether. The material was dissolved in hot iso-propanol, cooled to room temperature and re-precipitated with diethyl ether. The solid was washed with diethyl ether and dried under high vacuum to provide 2-(1-methylimidazol-2-yl)-3*H*-imidazo[4,5-*b*]pyridine-5-carboxylic acid methyl ester (Im-Ip-OMe) (240 mg, 50% Yield) as a tan

solid.  $^1\text{H NMR}$  (300 MHz,  $\text{DMSO-d}_6$ )  $\delta$  7.99 (s, 2 H), 7.51 (s, 1H), 7.20 (s, 1H), 4.18 (s, 3H), 3.88 (s, 3H);  $^{13}\text{C}$  (75 MHz,  $\text{DMSO-d}_6$ )  $\delta$  147.7, 142.0, 137.1, 131.8, 129.6, 126.9, 119.9, 109.3, 52.8, 35.8; HR-MS ( $\text{EI}^+$ ): calculated for  $\text{C}_{12}\text{H}_{11}\text{N}_5\text{O}_2$ : 257.0910; found: 257.0920.

#### 7.4.2.7 2-(1-Methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid (*Im-Ip-OH*) (**9**).

2-(1-Methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (*Im-Ip-OMe*) (250 mg, 0.97 mmol) dissolved in MeOH (2 ml) and KOH (4 N, 3 ml) was heated to 50 °C for 4 h. The methanol was removed in vacuo and the aqueous layer washed with EtOAc (2 x 10 mL) to remove any starting material and trace impurities. The pH of the aqueous layer was then adjusted slowly to pH = 4 with 1 N HCl upon which time a cloudy beige precipitate formed. The mixture was placed in a falcon tube and the precipitate concentrated by centrifugation. The supernatant was decanted and the solid dried under high vacuum to give 2-(1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid (**9**) (154 mg, 65% Yield) as a brown solid.  $^1\text{H NMR}$  (300 MHz,  $\text{DMSO-d}_6$ )  $\delta$  7.99 (s, 2H), 7.51 (d, 1H,  $J = 0.9$  Hz), 7.20 (d, 1H,  $J = 0.9$  Hz), 4.18 (s, 3H); HR-MS ( $\text{EI}^+$ ): calculated for  $\text{C}_{11}\text{H}_9\text{N}_5\text{O}_2$ : 243.0750; found: 243.0740.

### 7.4.3 Oligomer Synthesis

Oligomers were synthesized on solid support using Boc- $\beta$ -PAM resin (0.59 meq/g). Stepwise elongation of the oligomers was done according to previously published protocols.<sup>19</sup> The synthesis of compound **1** has been previously reported.<sup>14</sup>

#### 7.4.3.1 Preparation of Base Resin R- $\beta$ -Bi-Py-Bi-Py- $\gamma$ -NHBoc (**BRI**).

To a manual solid phase synthesis vessel was added Boc- $\beta$ -PAM resin (0.3 g). The resin was washed with DMF (15 mL) and allowed to swell for 15 min while shaking at room temperature. The resin was then washed with DCM (~30 mL), followed by 80% TFA in DCM (~30 mL) to remove the Boc-group. The resin was then agitated at room temperature in 80% TFA/DCM for another 25 min to provide the deprotected resin bound amine (R- $\beta$ -NH<sub>2</sub>). Following Boc-deprotection, the resin was washed with DCM and 10% DIEA in DMF to neutralize and prepare for coupling. Simultaneously, in a separate reaction vessel, Boc-Py-Bi-OH (189 mg, 531  $\mu\text{M}$ ), HBTU (191 mg, 504  $\mu\text{M}$ ), DIEA (137 mg, 185  $\mu\text{L}$ , 1.06 mM) and DMF (1.2 mL) was mixed and allowed to activate at room temperature for 25 min. This mixture was then added to the solid phase synthesis vessel containing R- $\beta$ -NH<sub>2</sub>. Coupling was allowed to proceed at room temperature with agitation for 3-6

h. Initial loading of the resin requires elongated coupling times. Following coupling, the resin was acylated by the addition of acetic anhydride to the mixture and shaking for 15 min. The addition of the next Boc-Py-Bi-OH (**25**) dimer was incorporated and deprotected as described above to provide the resin bound fragment (R- $\beta$ -Bi-Py-Bi-Py-NH<sub>2</sub>). To this fragment was added a preactivated mixture of Boc- $\gamma$ -OH (180 mg, 885  $\mu$ M), HBTU (319 mg, 841  $\mu$ M), DIEA (229 mg, 308  $\mu$ L, 1.77 mM). Coupling was allowed to proceed for 3 h at room temperature with agitation. The resin was then capped with acetic anhydride as described above to provide the base resin R- $\beta$ -Bi-Py-Bi-Py- $\gamma$ -NHBoc (**BR1**). **BR1** was then washed with DCM followed by MeOH and Et<sub>2</sub>O. The resin was then dried under high vacuum and stored for subsequent use.

#### 7.4.3.2 *Im-Im-Im-Im- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp (2)*.

**BR1** (50 mg) was added to a manual solid phase synthesis vessel. The resin was washed with DCM (~15 mL), followed by deprotection with 80% TFA in DCM. The resin was shaken at room temperature in the 80% TFA solution for 25 min. The resin was then drained, washed with DCM, and neutralized with 10% DIEA in DMF. A pre-activated mixture of Boc-Im-Im-OH (54 mg, 148  $\mu$ M), HBTU (53 mg, 140  $\mu$ M), DIEA (38 mg, 52  $\mu$ L, 295  $\mu$ M) and DMF (400  $\mu$ L) was then added to the reaction vessel and coupling was allowed to proceed for 3 h at room temperature, followed by capping with acetic anhydride as described for **BR1** to give R- $\beta$ -Bi-Py-Bi-Py- $\gamma$ -Im-Im-NHBoc. Following resin deprotection as described above, Im-Im-OH was activated as described for Boc-Im-Im-OH. Coupling of Im-Im-OH to the resin was allowed to proceed overnight at room temperature to provide R- $\beta$ -Bi-Py-Bi-Py- $\gamma$ -Im-Im-Im-Im. The resin was treated with the cleavage protocol outlined below to provide Im-Im-Im-Im- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp (**2**) in 5% yield. MALDI-TOF-MS: calculated for C<sub>58</sub>H<sub>66</sub>N<sub>23</sub>O<sub>8</sub>: 1212.55; found 1212.50 [M+H]<sup>+</sup>.

#### 7.4.3.3 *Im-Ip-Im-Im- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp (5)*.

**BR1** (50 mg) was added to a manual solid phase synthesis vessel and R- $\beta$ -Bi-Py-Bi-Py- $\gamma$ -Im-Im-NHBoc was prepared as described above for **2**. Following deprotection, washing and neutralization as described above, a pre-activated mixture of Im-Ip-OH (**14**) (21.5 mg, 88.5  $\mu$ M), HBTU (32 mg, 84  $\mu$ M), DIEA (23 mg, 31  $\mu$ L, 177  $\mu$ M), DMF (400  $\mu$ L) was added to the vessel containing R- $\beta$ -Bi-Py-Bi-Py- $\gamma$ -Im-Im-NH<sub>2</sub>. Coupling was allowed to proceed overnight at room temperature to provide R- $\beta$ -Bi-Py-Bi-Py- $\gamma$ -Im-Im-Ip-Im. The resin was treated with the cleavage protocol outlined below to provide Im-Ip-Im-Im- $\gamma$ -Py-Bi-Py-Bi- $\beta$ -Dp **5** in 2.2% yield. MALDI-TOF-MS: calculated

for  $C_{59}H_{64}N_{23}O_7$ ; 1206.54; found 1206.60 [M+H]<sup>+</sup>.

#### 7.4.3.4 *Im-Im-Im-Ip-γ-Py-Bi-Py-Bi-β-Dp (3)*.

**BR1** (50 mg, 0.81 meq/g) was added to a manual solid phase synthesis vessel. The resin was treated with the cleavage protocol outlined below to provide *Im-Im-Im-Ip-γ-Py-Bi-Py-Bi-β-Dp 3* in 3% yield. MALDI-TOF-MS: calculated for  $C_{59}H_{64}N_{23}O_7$ ; 1206.50; found 1206.50 [M+H]<sup>+</sup>.

#### 7.4.3.5 *Im-Im-Ip-Im-γ-Py-Bi-Py-Bi-β-Dp (4)*.

**BR1** (70 mg, 0.59 meq/g) was added to a manual solid phase synthesis vessel. The resin was washed with DCM (~15 mL), followed by deprotection with 80% TFA in DCM. The resin was shaken at room temperature in the 80% TFA solution for 25 min. The resin was then drained, washed with DCM, neutralized with 50% DIEA in DCM, and washed with DMF. A pre-activated mixture of Boc-Im-OH (50 mg, 207 μmol), HBTU (79 mg, 208 μmol), DIEA (53 mg, 72 μL, 413 μmol) and DMF (900 μL) was then added to the reaction vessel and coupling was allowed to proceed for 12 h at room temperature, followed by capping with acetic anhydride as described for **BR1** to give R-β-Bi-Py-Bi-Py-γ-Im-NHBoc. Following resin deprotection as described above, Boc-Im-Ip-OH (**12**) was activated as described for Boc-Im-OH. Coupling of **12** to the resin was allowed to proceed overnight at room temperature to provide R-β-Bi-Py-Bi-Py-γ-Im-Ip-Im-NHBoc. Following resin deprotection as described above, Im-CCl<sub>3</sub> (2-Trichloroacetyl-1-methylpyrrole) (47 mg, 207 μmol) and DIEA (53 mg, 72 μL, 413 μmol) were dissolved in NMP (900 μL) and added to the reaction vessel. Coupling of Im-CCl<sub>3</sub> to the resin was allowed to proceed overnight at 32 °C to provide R-β-Bi-Py-Bi-Py-γ-Im-Ip-Im-Im. The resin was treated with the cleavage protocol outlined below to provide *Im-Im-Ip-Im-γ-Py-Bi-Py-Bi-β-Dp 4* in 3% yield. MALDI-TOF-MS calculated for  $C_{59}H_{64}N_{23}O_7$ ; 1206.54; found 1206.50 [M+H]<sup>+</sup>.

#### 7.4.3.6 *ImIp-ImIp-γ-PyBi-PyBi-β-Dp (6)*.

**BR1** (50 mg) was added to a manual solid phase synthesis vessel. The resin was treated with the cleavage protocol outlined below to provide *Im-Ip-Im-Ip-γ-Py-Bi-Py-Bi-β-Dp 6* in 2% yield. MALDI-TOF-MS calculated for  $C_{60}H_{62}N_{23}O_6$ ; 1200.52; found 1200.50 [M+H]<sup>+</sup>.

#### 7.4.4 *Resin Cleavage Procedure*

A sample of resin (20-100 mg) was washed with DCM followed by the addition of

dimethylaminopropylamine (Dp) (1 mL). The mixture was heated to 80 °C for 2 h with occasional agitation. The resin was then filtered and washed with 0.1% TFA in water (7 mL). The combined filtrate was collected and subjected to purification by reverse phase preparatory HPLC using a Waters C<sub>18</sub> column and 0.1% TFA/ACN solvent system. Appropriate fractions from the HPLC purification were checked for purity by analytical HPLC and characterized by MALDI-TOF spectroscopy. Pure fractions were then pooled, flash frozen using liquid nitrogen and lyophilized to a dry solid for later use.

#### *7.4.5 Footprinting Experiments*

Plasmids pEF16 was constructed using standard methods. DNase I footprint titrations were performed according to standard protocols.<sup>17</sup>

#### *7.4.6 NF-κB Electrophoretic Mobility Shift Assay*

##### *7.4.6.1 Materials*

Jurkat Nuclear Extract containing activated NF-κB was purchased from Active Motif (36013) and diluted as necessary just prior to use with Buffer C (20 mM HEPES pH 7.9, 420 mM KCl, 1.5 mM MgCl<sub>2</sub>, 25% glycerol, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 1 μg/ml aprotinin). The final working concentration was 0.5-2.5 μg/μl. The identity of the NF-κB band was confirmed by antibody supershift (Figure 7.9) using antibodies against p50 (sc-7178, sc-114) and p65 (sc-372) from Santa Cruz Biotechnology.

##### *7.4.6.2 Antibody Supershift*

The NF-κB antibody supershift data is shown in Figure 7.9. NF-κB antibodies (sc-7178 and sc-114 against p50 and sc-372 against p65) came from Santa Cruz Biotechnology. The actin antibody was purchased from Sigma-Aldrich (A5441). Antibodies were diluted to 0.1 μg/μl just prior to use with ice cold PBS. Complete binding reactions contained 0.1 ng labeled probe, 10 mM Tris•HCl pH 7.5, 50 mM NaCl, 10% glycerol, 1% NP-40, 1 mM EDTA, 0.1 mg/ml poly(dI-dC), 1 μl antibody (0.1 μg), and 4 μg (2 μl) nuclear extract in a total volume of 10 μl. Nuclear extract was incubated with antibody at 4 °C for 1 hour followed by addition of binding buffer and probe. The complete reactions were allowed to incubate a further 30 min at room temperature, loaded onto pre-run 5% acrylamide, 5% glycerol gels, and resolved for 2 h at 150 volts. Gels were pre-run for 15-45 min prior to loading. The gel and running buffer was 24.8 mM Tris Base, 190 mM glycine, 1 mM 0.5

M EDTA.

#### 7.4.6.3 Sequence of Gel Shift Probes

Oligonucleotides were purchased from Integrated DNA Technologies. The size of both probes following labeling was 40 basepairs. The match probe contained an imbedded  $\kappa$ B site from the intronic enhancer of the immunoglobulin  $\kappa$  light-chain gene (underlined below). As a control, the sequence of the mismatch probe contained a mutated  $\kappa$ B that prevented NF- $\kappa$ B binding (changes bolded below).

*Match Probe:*

5' AATTCATGCAGTTGAGGGGACTTTCCAGGCATGCA 3'

*Match Complementary Sequence:*

5' AGCTTGCATGCCTGGGAAAGTCCCCTCAACTGCATG 3'

*Mismatch Probe:*

5' AATTCATGCAGTTGAGGTTACTTTCCAGGCATGCA 3'

*Mismatch Complementary Sequence:*

5' AGCTTGCATGCCTGGGAAAGTAACTCAACTGCATG 3'

#### 7.4.6.4 Preparation of 3' Labeled Probes

In separate reactions, 1.1 pmol of match or mismatch oligonucleotide was annealed to an equal amount of its complementary sequence by heating at 95 °C for 1 minute and cooling slowly to room temperature. The 3' ends were radiolabeled with  $^{32}$ P using Sequenase Version 2.0 (Amersham) and  $\alpha$ - $^{32}$ P dATP (PerkinElmer). The resulting labeled probes were purified with G-50 Microspin Columns (Amersham). Probe concentration was estimated at 0.8 ng/ $\mu$ l assuming 100% recovery based on a calculated molecular weight of 24,662 g/mol. Just prior to use, labeled probes were diluted in water to 0.1 ng/ $\mu$ l.

#### 7.4.6.5 Gelshift Screen

Dry HPLC purified aliquots of polyamide were dissolved in water and their concentration determined by measuring the absorbance at 310 nm ( $\epsilon = 69,520 \text{ M}^{-1} \text{ C}^{-1}$ ). 10x solutions (1  $\mu$ M and 100 nM) of each polyamide were prepared from serial dilutions of the concentrated stock. Binding reactions contained 0.1 ng labeled probe, 10 mM Tris•HCl pH 7.5, 50 mM NaCl, 10% glycerol, 1% NP-40, 1 mM EDTA, 0.1 mg/ml poly(dI-dC), 1  $\mu$ l 10x polyamide (final concentration 10 nM or 100 nM),

and 3  $\mu\text{g}$  (1-2  $\mu\text{l}$ ) nuclear extract in a total volume of 10  $\mu\text{l}$ . Reactions were allowed to equilibrate at room temperature without nuclear extract for 3 h, followed by addition of protein. After a further 30 min at room temperature, the reactions were loaded and resolved on non-denaturing 5% acrylamide, 5% glycerol gels for 2-2.5 h at 150 volts. The gels were immediately dried and exposed to a phosphoimage storage plate for at least 8 h. The gel and running buffer was 24.8 mM Tris Base, 190 mM glycine, 1 mM 0.5 M EDTA. Gels were pre-run for 15 to 45 min prior to loading.

#### 7.4.6.6 Gelshift Titration

Reactions were prepared as described above with a constant 0.1 ng probe and 3  $\mu\text{g}$  nuclear extract, except that 10x solutions of polyamide were prepared to give a final concentration range of 500 pM to 500 nM.

#### 7.4.6.7 Data Analysis

A Typhoon 8600 Variable Mode Imager was used to visualize the gels and band intensity was quantified using ImageQuant software version 5.1 from Molecular Dynamics. The fraction of DNA bound to NF- $\kappa$ B was calculated as the intensity ratio of the NF- $\kappa$ B shifted band to the sum total of all bands, shifted and unshifted.  $\text{EC}_{50}$  values were determined graphically as the polyamide concentration required to reduce NF- $\kappa$ B band intensity to half its value. The data was plotted vs. polyamide concentration and fit to the Hill equation ( $n=1$ ) using Kaleidagraph software.

## 7.5 Notes and References

1. (a) Olenyuk, B. Z.; Zhang, G.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G. Jr.; Dervan, P. B. Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16768-16773. (b) Darnell, J. E. Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* **2002**, *2*, 740-748. (c) Pandolfi, P. P. Transcription therapy for cancer. *Oncogene* **2001**, *20*, 3116-3127. (d) Dervan, P. B.; Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.
2. (a) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. Netropsin, A new antibiotic produced by a streptomycetes. *J. Am. Chem. Soc.* **1951**, *73*, 341-343. (b) Arcamone, F. N. V.; Penco, S.; Orezzi, P.; Nicoletta, V.; Pirelli, A. Structure and synthesis of Distamycin A. *Nature* **1964**, *203*, 1064-1065.
3. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376-1380. (b) Pelton, J. G.; Wemmer, D. E. Structural characterization of a 2: 1 distamycin Ad(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723-5727. (c) Wemmer, D. E. Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 439-461. (d) Buchmueller, K. L.; Staples, A. M.; Howard, C.

M.; Horick, S. M.; Uthe, P. B.; Minh Le, N.; Cox, K. K.; Nguyen, B.; Pacheco, K. A. O.; Wilson, D. W.; Lee, M. Extending the language of DNA molecular recognition by polyamides: unexpected influence of imidazole and pyrrole arrangement on binding affinity and specificity. *J. Am. Chem. Soc.* **2004**, *127*, 742-750. (e) Baraldi, P. G.; Bovero, A.; Fruttarolo, F.; Preti, D.; Tabrizi, M. A.; Pavani, M. G.; Romagnoli, R. DNA minor groove binders as potential antitumor and antimicrobial agents. *Med. Res. Rev.* **2004**, *24*, 475-528. (f) Reddy, P. M.; Toporowski, J. W.; Kahane, A. L.; Bruce, T. C. Recognition of a 10 base pair sequence of DNA and stereochemical control of the binding affinity of chiral hairpin polyamide-Hoechst 33258 conjugates. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5531-5536.

4. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **1998**, *391*, 468-471.

5. (a) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science* **1998**, *282*, 111-115. (b) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. Structural effects of DNA sequence on TA recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **2000**, *295*, 557-567.

6. (a) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Sequence selectivity of 3-hydroxypyrrole/pyrrole ring pairings in the DNA minor groove. *J. Am. Chem. Soc.* **1999**, *121*, 11621-11629. (b) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. Affinity and specificity of multiple hydroxypyrrole/pyrrole ring pairings for coded recognition of DNA. *J. Am. Chem. Soc.* **1999**, *121*, 260-261.

7. (a) Hays, F. A.; Teegarden, A.; Jones, Z. J. R.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Shing Ho, P. How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7157-7162. (b) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham 3rd, T. E.; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.* **2004**, *87*, 3799-3813. (c) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham 3rd, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.* **2005**, *89*, 3721-3740. (d) Wu, H.; Crothers, D. M. The locus of sequence-directed and protein-induced DNA bending. *Nature* **1984**, *308*, 509-513. (e) Steitz, T. A. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* **1990**, *23*, 205-280. (f) Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2930-2934. (g) Paoletta, D. N.; Palmer, R.; Schepartz, A. DNA targets for certain bZIP proteins distinguished by an intrinsic bend. *Science* **1994**, *264*, 1130-1133. (h) Kahn, J. D.; Yun, E.; Crothers, D. M. Detection of localized DNA flexibility. *Nature* **1994**, *368*, 163-166. (i) Geierstanger, B. H.; Wemmer, D. E. Complexes of the minor groove of DNA. *Annu. Rev. Biophys. Biochem. Struct.* **1995**, *24*, 463-493. (j) Hansen, M. R.; Hurley, L. H. Pluramycins. Old drugs having modern friends in structural biology. *Acc. Chem. Res.* **1996**, *29*, 249-258. (k) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226. (l) Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B. Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor

groove of DNA. *Helv. Chim. Acta* **2002**, *85*, 4485-4517.

8. Briehn, C. A.; Weyermann, P.; Dervan, P. B. Alternative heterocycles for DNA recognition: the benzimidazole/imidazole pair. *Chem. Eur. J.* **2003**, *9*, 2110-2112.

9. (a) Renneberg, D.; Dervan, P. B. Imidazopyridine/Pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc.* **2003**, *125*, 5707-5716. (b) Marques, M. A.; Doss, R. M.; Foister, S.; Dervan, P. B. Expanding the repertoire of heterocycle ring pairs for programmable minor groove DNA recognition. *J. Am. Chem. Soc.* **2004**, *126*, 10339-10349.

10. Doss, R. M.; Marques, M. A.; Foister, S.; Chenoweth, D. M.; Dervan, P. B. Programmable oligomers for minor groove DNA recognition. *J. Am. Chem. Soc.* **2006**, *128*, 9074-9079.

11. Viger, A.; Dervan, P. B. Exploring the limits of benzimidazole DNA-binding oligomers for the hypoxia inducible factor (HIF) site. *Bioorg. Med. Chem.* **2006**, *14*, 8539-8549.

12. (a) Natoli, G.; Saccani, S.; Bosisio, D.; Marazzi, I. Interactions of NF-kappaB with chromatin: the art of being at the right place at the right time. *Nat. Immunol.* **2005**, *6*, 439-445. (b) Karin, M.; Greten, F. R. NF-kappaB: linking inflammation and immunity to cancer development and progression. *Nature Rev. Immunol.* **2005**, *5*, 749-759. (c) Aggarwal, B. B. Nuclear factor-kappaB: the enemy within. *Cancer Cell* **2004**, *6*, 203-208. (d) Pande, V.; Ramos, M. NF-kappaB in human disease: current inhibitors and prospects for de novo structure based design of inhibitors. *Curr. Med. Chem.* **2005**, *12*, 357-374.

13. Chen, F. E.; Huang, D. B.; Chen, Y. Q.; Ghosh, G. Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA. *Nature* **1998**, *391*, (6665), 410-413.

14. (a) Swalley, S. E.; Baird, E. E.; Dervan, P. B. Recognition of a 5'-(A,T)GGG(A,T)<sub>2</sub>-3' sequence in the minor groove of DNA by an eight-ring hairpin polyamide. *J. Am. Chem. Soc.* **1996**, *118*, 8198-8206. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. Discrimination of 5'-GGGG-3', 5'-GCGC-3', and 5'-GGCC-3' sequences in the minor-groove of DNA by 8-ring hairpin polyamides. *J. Am. Chem. Soc.* **1997**, *119*, 6953-6961.

15. Wurtz, N. R.; Pomerantz, J. L.; Baltimore, D.; Dervan, P. B. Inhibition of DNA binding by NF-kappa B with pyrrole-imidazole polyamides. *Biochemistry* **2002**, *41*, 7604-7609.

16. (a) Briehn, C. A.; Weyermann, P.; Dervan, P. B. Alternative heterocycles for DNA recognition: the benzimidazole/imidazole pair. *Chem. Eur. J.* **2003**, *9*, 2110-2112. (b) Renneberg, D.; Dervan, P. B. Imidazopyridine/Pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc.* **2003**, *125*, 5707-5716. (c) Marques, M. A.; Doss, R. M.; Foister, S.; Dervan, P. B. Expanding the repertoire of heterocycle ring pairs for programmable minor groove DNA recognition. *J. Am. Chem. Soc.* **2004**, *126*, 10339-10349. (d) Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B. Shape selective recognition of TA base pairs by hairpin polyamides containing N-terminal 3-methoxy (and 3-chloro) thiophene residues. *Bioorg. Med. Chem.* **2003**, *11*, 4333-4340.

17. Trauger, J. W.; Dervan, P. B. Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. *Methods Enzymol.* **2001**, *340*, 450-466.

18. Austin, M. W.; Blackburn, J. R.; Ridd, J. H.; Smith, B. V. The kinetics and mechanism of heteroaromatic nitration. Part II. Pyrazole and imidazole. *J. Chem. Soc.* **1965**, 1051-1057.

19. Baird, E. E.; Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

20. (a) Qu, X. G.; Ren, J. S.; Riccelli, P. V.; Benight, A. S.; Chaires, J. B. Enthalpy/entropy

compensation: influence of DNA flanking sequence on the binding of 7-amino actinomycin D to its primary binding site in short DNA duplexes. *Biochemistry* **2003**, *42*, 11960-11967. (b) Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B. Structure of a beta-alanine-linked polyamide bound to a full helical turn of purine tract DNA in the 1: 1 motif. *J. Mol. Biol.* **2002**, *320*, (1), 55-71. (c) Melander, C.; Herman, D. M.; Dervan, P. B. Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif. *Chem. Eur. J.* **2000**, *6*, 4487-4497.

21. Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. Specific binding of Hoechst 33258 to the d(CGCAAATTTGCG)<sub>2</sub> duplex: calorimetric and spectroscopic studies. *J. Mol. Biol.* **1997**, *271*, 244-257.

22. McKay, S. L.; Haptonstall, B.; Gellman, S. H. Beyond the hydrophobic effect: attractions involving heteroaromatic rings in aqueous solution. *J. Am. Chem. Soc.* **2001**, *123*, 1244-1245.

23. Ji, Y. H.; Bur, D.; Hasler, W.; Schmitt, V. R.; Dorn, A.; Bailly, C.; Waring, M. J.; Hochstrasser, R.; Leupin, W. Tris-benzimidazole derivatives: design, synthesis and DNA sequence recognition. *Bioorg. Med. Chem.* **2001**, *9*, 2905-2919.

24. Tisne, C.; Hartmann, B.; Delepierre, M. NF-kappa B binding mechanism: a nuclear magnetic resonance and modeling study of a GGG-->CTC mutation. *Biochemistry* **1999**, *38*, 3883-3894.

25. Huang, D.; Phelps, C. B.; Fusco, A. J.; Ghosh, G. Crystal structure of a free kappaB DNA: insights into DNA recognition by transcription factor NF-kappaB. *J. Mol. Biol.* **2005**, *346*, 147-160.

## Chapter 8: Fluorescent Sequence-Specific dsDNA Binding Oligomers

*The text of this chapter was taken in part from a manuscript coauthored with Anne Viger and Peter B. Dervan\* (Caltech)*

(Chenoweth, D.M., Viger, A., Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2216-2217.)

**Abstract**

Sequence-specific detection methods for double-stranded DNA that obviate the need for denaturation would provide a powerful tool for bioorganic chemistry and genetics. As part of a sustained effort to develop sequence-specific fluorescent DNA detection methods, two programmable oligomers have been synthesized which target their respective sequences 5'-WTACGW-3' and 5'-WGGGGW-3' (W = A or T). The two oligomers were found to fluoresce weakly in the absence of DNA but showed significant fluorescence enhancement by the addition of match DNA. The fluorescence is shown to increase in a concentration-dependent manner, and the intensity varies depending on the number of mismatch sites incorporated into the DNA hairpins. This new class of oligomers provides a method to detect DNA sequences without denaturation and in the absence of conjugation to a dye molecule. This is a first step toward sequence-specific DNA-binding molecules containing a fluorescent switch integrated as part of the recognition modules.

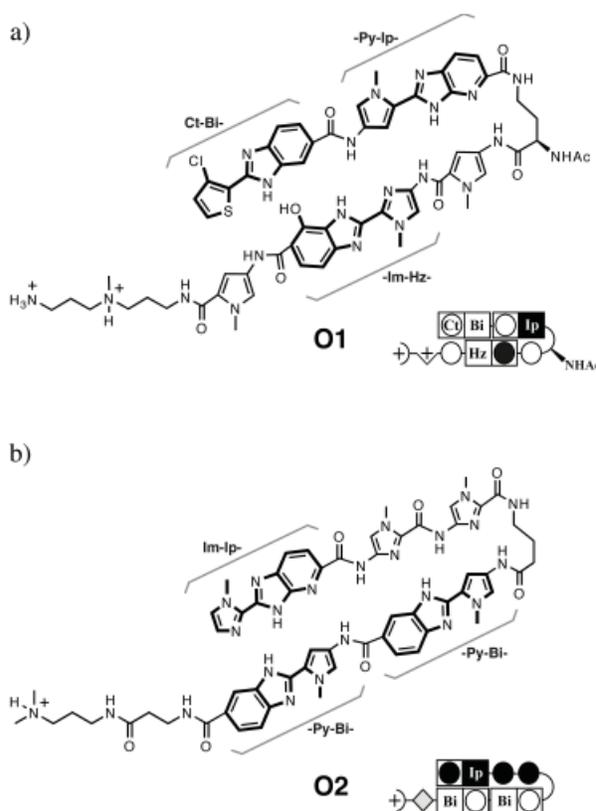
## 8.1 Introduction

Sequence-specific detection methods for double-stranded DNA (dsDNA) that obviate the need for denaturation would provide useful tools for bioorganic chemistry and genetics.<sup>1</sup> Previous efforts, such as molecular beacons<sup>2</sup> or peptide nucleic acid-thiazole orange (PNA-TO) conjugates,<sup>3</sup> require harsh denaturation conditions for hybridization to single-stranded DNA.<sup>2-4</sup> Previous efforts from our laboratory for the sequence-specific detection of dsDNA have focused on pyrrole-imidazole (Py-Im) polyamide-fluorophore conjugates, such as tetramethylrhodamine (TMR) or thiazole orange (TO), that bind in the minor groove of DNA.<sup>5-7</sup> TMR fluorescence was shown to be quenched when the fluorophore was covalently linked to the ring nitrogen of a pyrrole recognition element within a polyamide.<sup>5</sup> Remarkably, fluorescence was restored in a sequencedependent manner upon binding to dsDNA.<sup>5</sup> Similarly, polyamide-TO intercalator conjugates also demonstrate fluorescence enhancement in the presence of match dsDNA.<sup>6</sup>

Having established Py-Im polyamide-dye conjugates as a suitable platform for sequence-specific fluorescent dsDNA detection,<sup>5,6</sup> we sought to develop a new class of fluorescent DNA binders wherein the fluorescent moiety is an integrated part of the recognition modules. We report here the design of sequence-specific fluorescent dsDNA-binding oligomers (Figure 8.1) which incorporate multiple 6-5 fused dimer recognition modules<sup>8</sup> and show a marked fluorescent enhancement upon excitation at 340 nm in the presence of dsDNA.

## 8.2 Results and Discussion

Oligomer **O1** contains the chlorothiophene-benzimidazole (Ct-Bi-), pyrrole-imidazopyridine (-Py-Ip-), and imidazolehydroxybenzimidazole (-Im-Hz-) recognition modules, whereas oligomer **O2** contains imidazole-imidazopyridine (Im-Ip-) and two pyrrole-benzimidazole (-Py-Bi-) recognition modules.

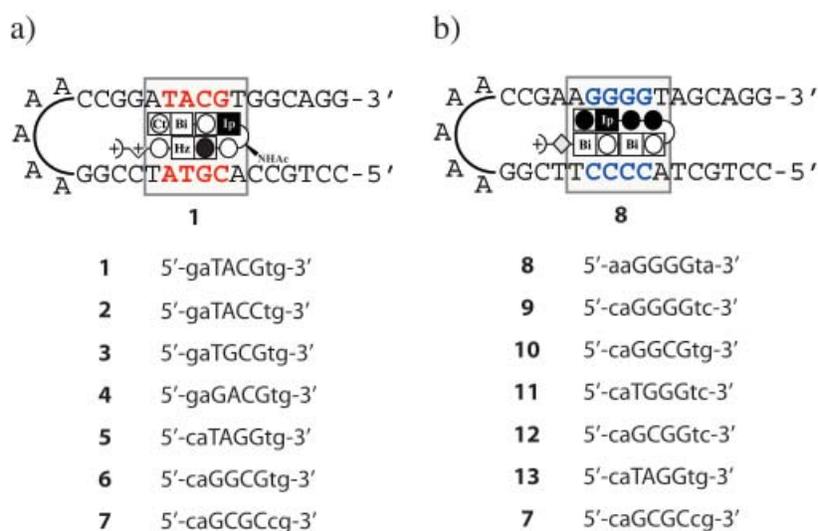


**Figure 8.1** Structure of oligomers. a) Oligomer **O1** containing Ct-Bi-, -Py-Ip-, and -Im-Hz- recognition modules. b) Oligomer **O2** containing Im-Ip- and two -Py-Bi- recognition modules.

modules. The binding affinities of **O1** and **O2** targeted to two biologically important sequences, 5'-ATACGT-3' (**O1**) and 5'-WGGGGW-3' (**O2**), were determined to be  $K_a = 1.6 \times 10^9 \text{ M}^{-1}$  and  $2.6 \times 10^9 \text{ M}^{-1}$ , respectively, by quantitative DNase I footprinting.<sup>9-11</sup>

A library of dsDNA hairpins containing six base-pair match and mismatch binding sites for **O1** and **O2** was used to investigate their emission properties (Figure 8.2). The dsDNA library for **O1** and **O2** contained match sites (**1** and **8**, respectively), single base-pair (bp) mismatch sites (**2–5** and **10–12**, respectively), double bp mismatch sites (**6** and **13**, respectively), and full mismatch sites (**7** for both oligomers). The dsDNA **9** contains the 4-G match site of oligomer **O2**; however, the flanking sequence has been changed to emphasize the effect on binding. The presence of G•C bp under the tail is expected to lower the binding affinity of **O2** as compared to that of dsDNA **8**.<sup>12</sup>

Oligomers **O1** and **O2** (1  $\mu\text{M}$  concentration) were each incubated with an increasing concentration (1 nM to 1  $\mu\text{M}$ ) of dsDNA, and their emission spectra were recorded after excitation at 340 nm. The oligomers



**Figure 8.2** Design of dsDNA library. a) dsDNA sequences used for **O1**. b) dsDNA sequences used for **O2**.

exhibited a marked increase in fluorescence upon addition of dsDNA containing their match site **1** and **8**, respectively (Figures 8.3 and 8.4).<sup>13</sup> Oligomer **O1** showed a moderate decrease in fluorescence intensity in the presence of dsDNA **2**, but proved to be much more sensitive to the incorporation of single base-pair mismatches at the alternate positions in dsDNAs **3–5** (Figure 8.4a). The incorporation of multiple base-pair mismatches in dsDNAs **6** and **7** showed a significant reduction in fluorescence intensity for **O1**. Oligomer **O2** exhibited a similar trend in sequence specificity, with a moderate decrease in fluorescence intensity observed upon incorporation of single base-pair mismatches (**9–12**) and a more significant decrease with multiple mismatches (**13** and **7**, Figure 8.4b).

### 8.3 Conclusion

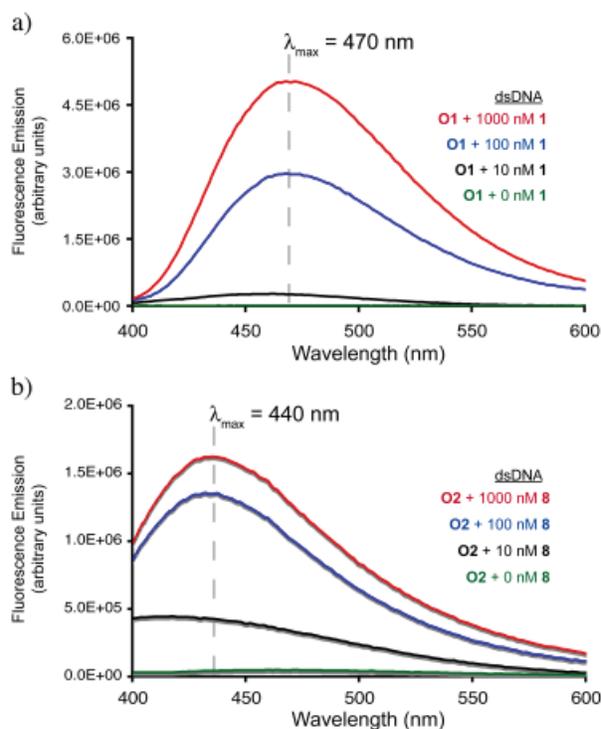
Sequence-specific DNA binding molecules containing a fluorescent switch integrated as part of the recognition modules provides a method to detect DNA sequences without denaturation and in the absence of conjugation to a dye molecule. Fluorescent oligomers may be useful as site-specific chromosome paints for telomeric and centromeric repeats<sup>14</sup> and could provide insight into cellular trafficking of DNA binding compounds.

### 8.4 Experimental

Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-HCl was purchased from United States Biochemical. All reagents were used without further purification. Water (18 M $\Omega$ ) was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2  $\mu$ m filtered. Reagent-grade chemicals were used as received, unless otherwise stated. Oligomers **O1** and **O2** were prepared by literature procedures.<sup>9,10</sup>

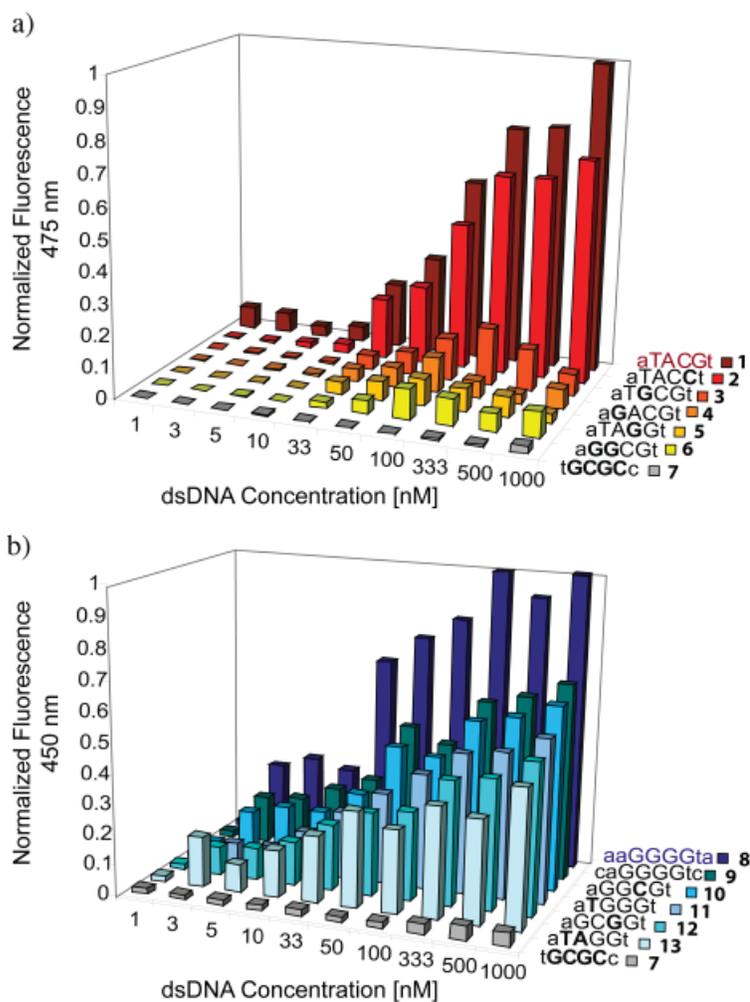
UV spectra were measured on a Agilent model 8453 diode-array spectrophotometer (Figure 8.5). Fluorescence spectra were measured with a Jobin Yvon/SPEX Fluorolog spectrofluorimeter (Model FL3-11) equipped with a Hamamatsu R928 PMT. Samples were excited at 340 nm using 8 nm emission and excitation slits and luminescence was observed from 400 to 600 nm at room temperature.

All measurements were performed in TKMC buffer [10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>]. The concentration of **O1** or **O2** was 1  $\mu$ M and the volume of solution used was 50  $\mu$ L for fluorescence measurements. The extinction coefficients for **O1** and **O2** were  $\epsilon = 58,700$  cm L mol<sup>-1</sup> at 330 nm and  $\epsilon = 69,200$  cm L mol<sup>-1</sup> at 340 nm respectively. The



**Figure 8.3** Fluorescence emission spectra of **O1** and **O2** (1  $\mu$ M) after 12 h incubation with their match binding site dsDNA ( $\lambda_{\text{Ex}}$  340 nm). a) Data for compound **O1**. b) Data for compound **O2**. The emission was shown to plateau beyond 1 equiv DNA. (See Section 8.6 Spectra and Supplemental Information for plots.)

concentration of hairpin DNA was varied from 1 nM to 1  $\mu$ M. Solutions containing **O1** and **O2** in the presence of varying concentrations of hairpin DNA were allowed to equilibrate for 12 h prior to fluorescence measurements.



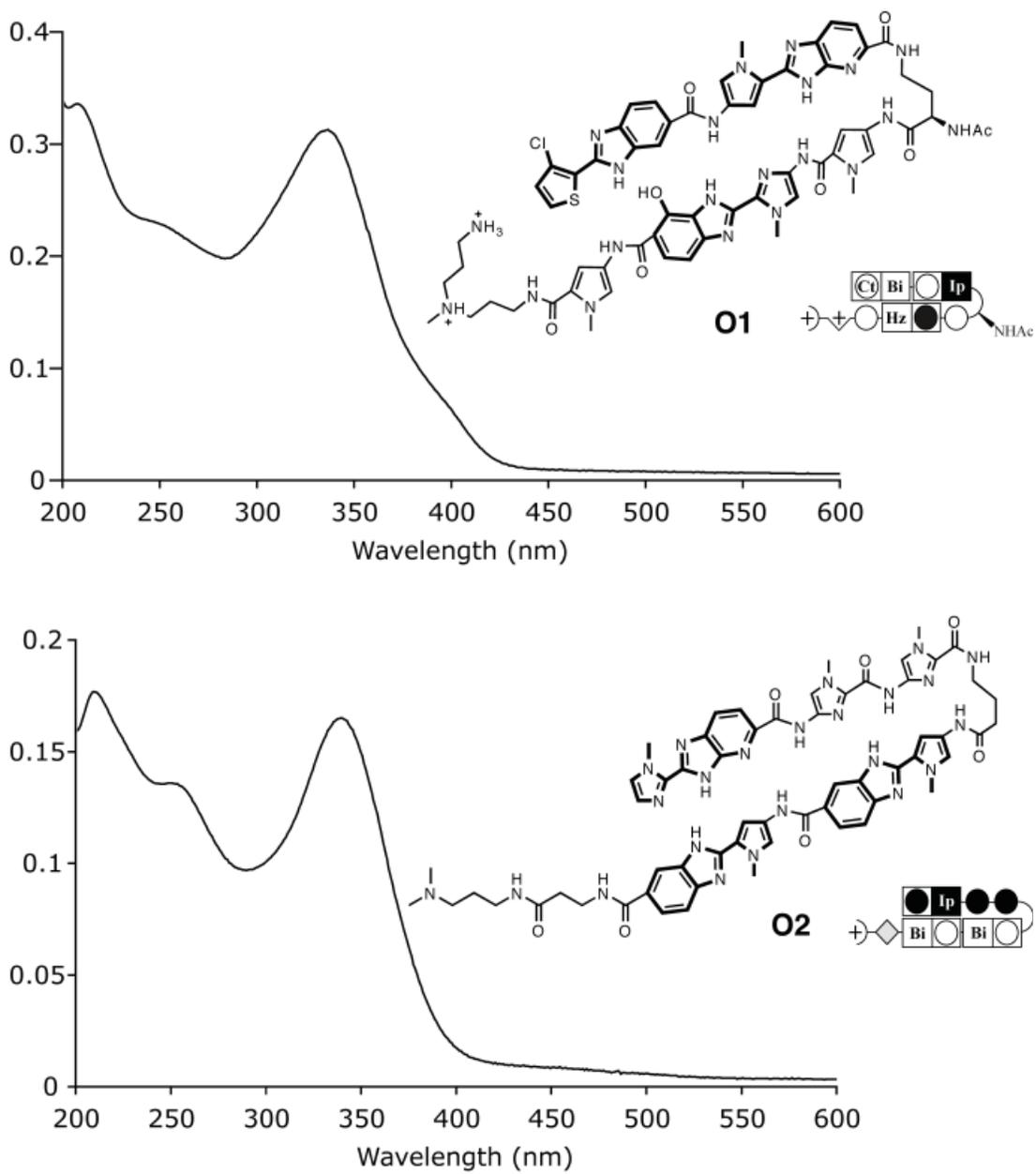
**Figure 8.4** Plot of dsDNA concentration versus normalized fluorescence for each dsDNA. a) Data for compound **O1**. b) Data for compound **O2**.

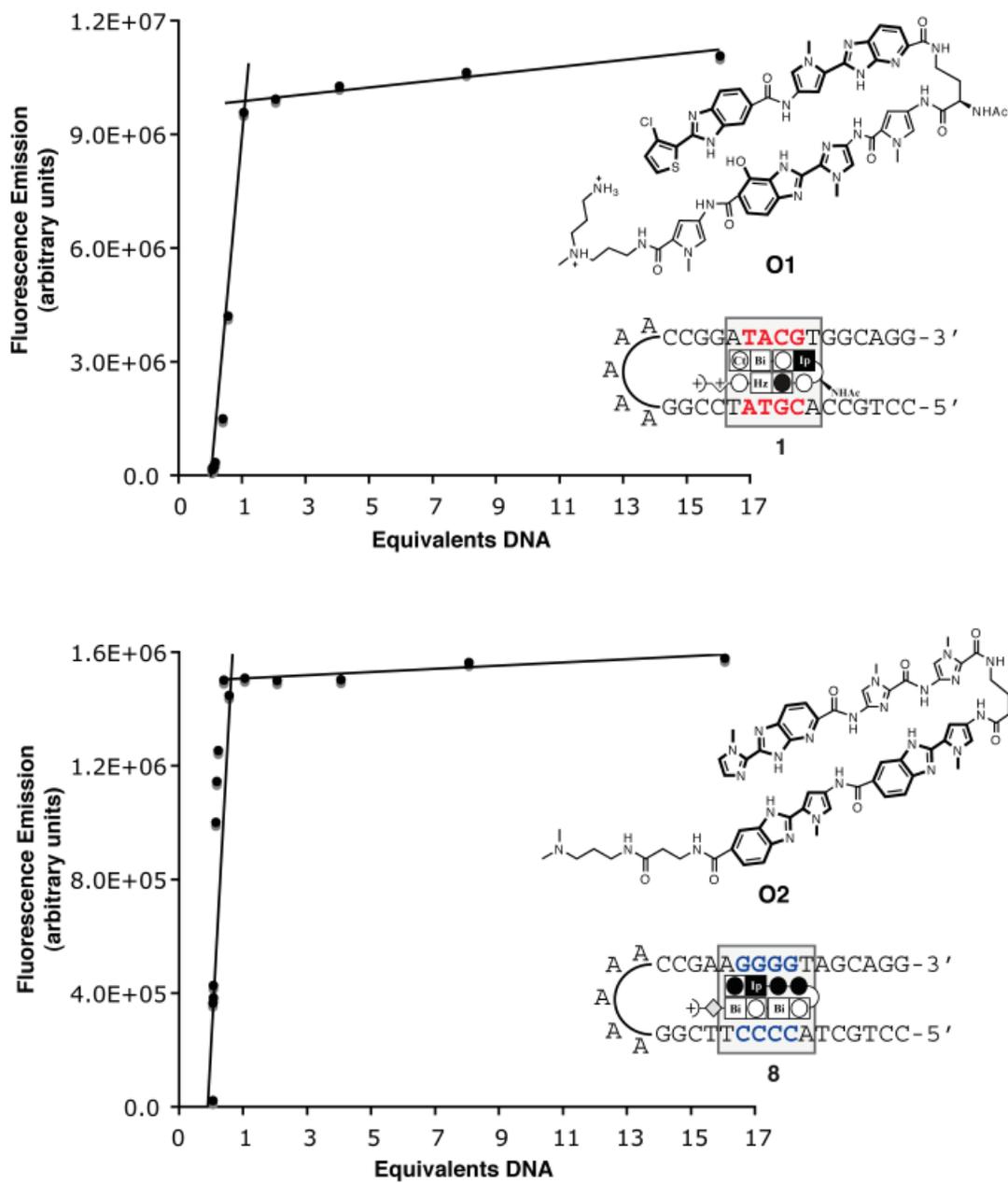
### 8.5 Notes and References

1. Chan, E. Y.; Goncalves, N. M.; Haeusler, R. A.; Hatch, A. J.; Larson, J. W.; Maletta, A. M.; Yantz, G. R.; Carstea, E. D.; Fuchs, M.; Wong, G. G.; Sullans, S. R.; Gilmanshin, R. DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags. *Genome Res.* **2004**, *14*, 1137-1146.
2. (a) Tyagi, S.; Kramer, F. R. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* **1996**, *14*, 303-308. (b) Tyagi, S.; Bratu, D. P.; Kramer, F. R. Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol.* **1998**, *16*, 49-53. (c) Kostrikis, L. G.; Tyagi, S.; Mhlanga, M. M.; Ho, D. D.; Kramer, F. R. Spectral genotyping of human alleles. *Science* **1998**, *279*, 1228-1229.
3. Svanvik, N.; Westman, G.; Wang, D.; Kubista, M. Light-up probes: thiazole orange-conjugated peptide nucleic acid for detection of target nucleic acid in homogeneous solution. *Anal. Biochem.* **2000**, *281*, 26-35.

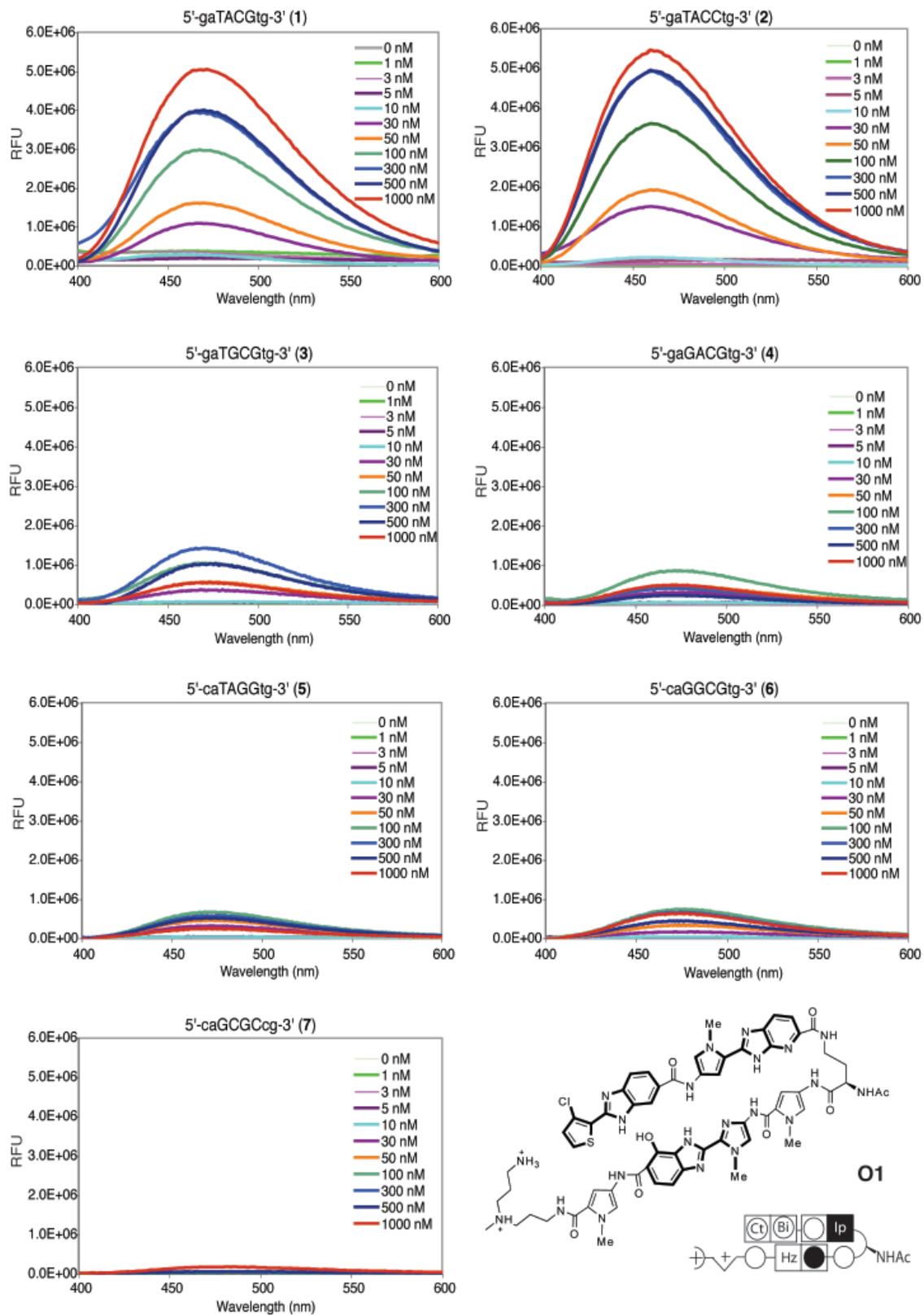
4. (a) Whitcombe, D.; Theaker, J.; Guy, S. P.; Brown, T.; Little, S. Detection of PCR products using self-probing amplicons and fluorescence. *Nat. Biotechnol.* **1999**, *17*, 804-807. (b) Thelwell, N.; Millington, S.; Solinas, A.; Booth, J.; Brown, T. Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res.* **2000**, *28*, 3752-3761. (c) Jenkins, Y.; Barton, J. K. A sequence-specific molecular light switch: tethering of an oligonucleotide to a dipyridophenazine complex of ruthenium (II). *J. Am. Chem. Soc.* **1992**, *114*, 8736-8738.
5. (a) Rucker, V. C.; Foister, S.; Melander, C.; Dervan, P. B. Sequence specific fluorescence detection of double strand DNA. *J. Am. Chem. Soc.* **2003**, *125*, 1195-1202. (b) Rucker, V. C.; Dunn, A. R.; Sharma, S.; Dervan, P. B.; Gray, H. B. Mechanism of sequence-specific fluorescent detection of DNA by *N*-methyl-imidazole, *N*-methyl-pyrrole, and alanine linked polyamides. *J. Phys. Chem. B* **2004**, *108*, 7490-7494.
6. Fechter, E. J.; Olenyuk, B.; Dervan, P. B. Sequence-specific fluorescence detection of DNA by polyamide-thiazole orange conjugates. *J. Am. Chem. Soc.* **2005**, *127*, 16685-16691.
7. Approaches based on peptide-thiazole orange conjugates show enhanced fluorescence in the presence of calf thymus (CT) DNA; however, DNA sequence specificity has not been observed. Carreon, J. R.; Mahon, K. P., Jr.; Kelley, S. O. Thiazole orange-peptide conjugates: sensitivity of DNA binding to chemical structure. *Org. Lett.* **2004**, *6*, 517-519.
8. Doss, R.; Marques, M. A.; Foister, S.; Chenoweth, D. M.; Dervan, P. B. Programmable oligomers for minor groove DNA recognition. *J. Am. Chem. Soc.* **2006**, *128*, 9074-9079.
9. Viger, A.; Dervan, P. B. Exploring the limits of benzimidazole DNA-binding oligomers for the hypoxia inducible factor (HIF) site. *Bioorg. Med. Chem.* **2006**, *14*, 8539-8549.
10. Chenoweth, D. M.; Poposki, J. A.; Marques, M. A.; Dervan, P. B. Programmable oligomers targeting 5'-GGGG-3' in the minor groove of DNA and NF-kappaB binding inhibition. *Bioorg. Med. Chem.* **2007**, *15*, 759-770.
11. The sequence 5'-ATACGT-3' is the hypoxia response element in the VEGF promotor and 5'-WGGGGW-3' is a sequence within the NF-κB response element.<sup>9,10</sup>
12. For a study of flanking sequence effects see: Swalley, S. E.; Baird, E. E.; Dervan, P. B. Effects of gamma-turn and beta-tail amino acids on sequence-specific recognition of DNA by hairpin polyamides. *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.
13. Quenching in the absence of DNA could be attributed to intramolecular quenching or solvent assisted quenching. Studies are underway to elucidate the mechanism.
14. (a) Gygi, M. P.; Ferguson, M. D.; Mefford, H. C.; Lund, K. P.; O'Day, C.; Zhou, P.; Friedman, C.; Engh, G.; Stolowitz, M. L.; Trask, B. J. Use of fluorescent sequence-specific polyamides to discriminate human chromosomes by microscopy and flow cytometry. *Nucleic Acids Res.* **2002**, *30*, 2790-2799. (b) Blattes, R.; Monod, C.; Susbielle, G.; Cuvier, O.; Wu, J.; Hsieh, T.; Laemmli, U. K.; Kas, E. Displacement of D1, HP1 and topoisomerase II from satellite heterochromatin by a specific polyamide. *EMBO J.* **2006**, *25*, 2397-2408.

## 8.6 Spectra and Supplemental Information

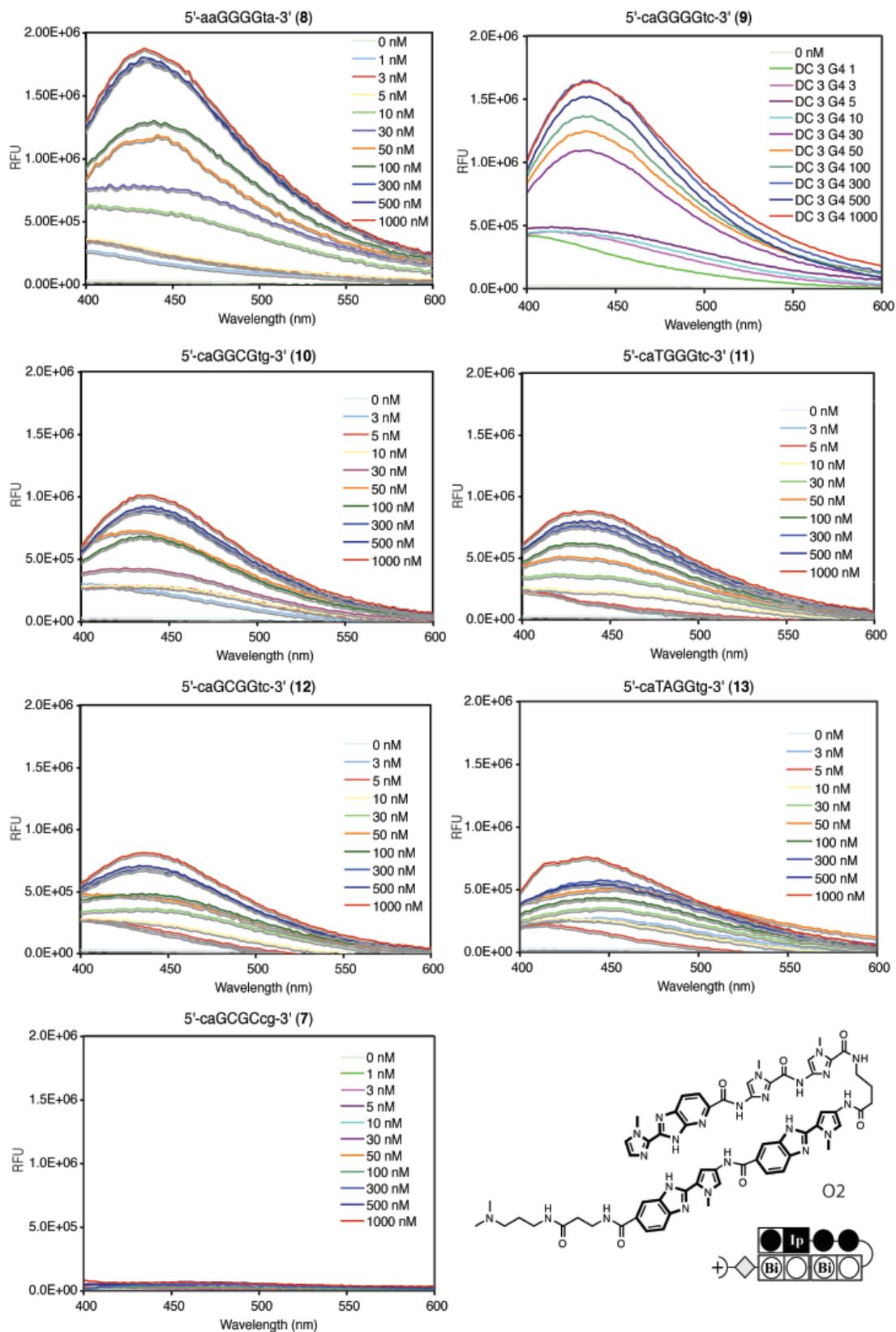
**Figure 8.5** UV spectra for compounds **O1** and **O2**.



**Figure 8.6** Fluorescence emission as a function of equivalents of DNA for compounds **O1** and **O2**.



**Figure 8.7** Fluorescence emission spectra compound **O1**.



**Figure 8.8** Fluorescence emission spectra compound **O2**.

## **Chapter 9: Polyamide/NCP Ligation and Profluorescent Azido- Carbostyrils**

*The research in this chapter on the NCP templated ligation of azide and alkyne containing polyamides was done in collaboration with Justin D. Cohen (Caltech).*

**Abstract**

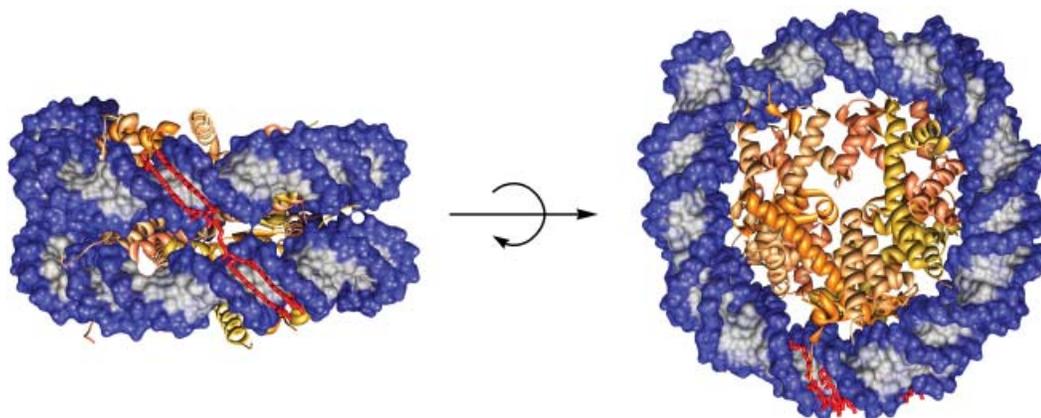
The nucleosome core particle (NCP) represents a physiologically relevant fundamental repeating unit of chromatin. The ability to modulate the structure of chromatin using sequence specific DNA binding small molecules could have potential applications in gene regulation, biotechnology, physical biology, and macromolecular structure studies. The possibility of self-assembling small molecules inside living cells using biological architectures represents a novel approach to molecular recognition. Small molecule probes may be used to report on dynamic information, regulate structural changes, and provide stabilization for structure determination of inherently dynamic macromolecules. Sequence-specific polyamides are a modular platform and could be used as DNA-binding building blocks to template the self-assembly of larger polyamides that target higher order structures such as the NCP, the tetra-NCP, the chromatin fiber, or other diverse nanometer scale structures inside the cell. This study provides a proof-of-principle experiment demonstrating that polyamides can be self-assembled using a high-order nucleoprotein structure (NCP) as a template.

## 9.1 Introduction

### 9.1.1 Templated Dimerization of Polyamides

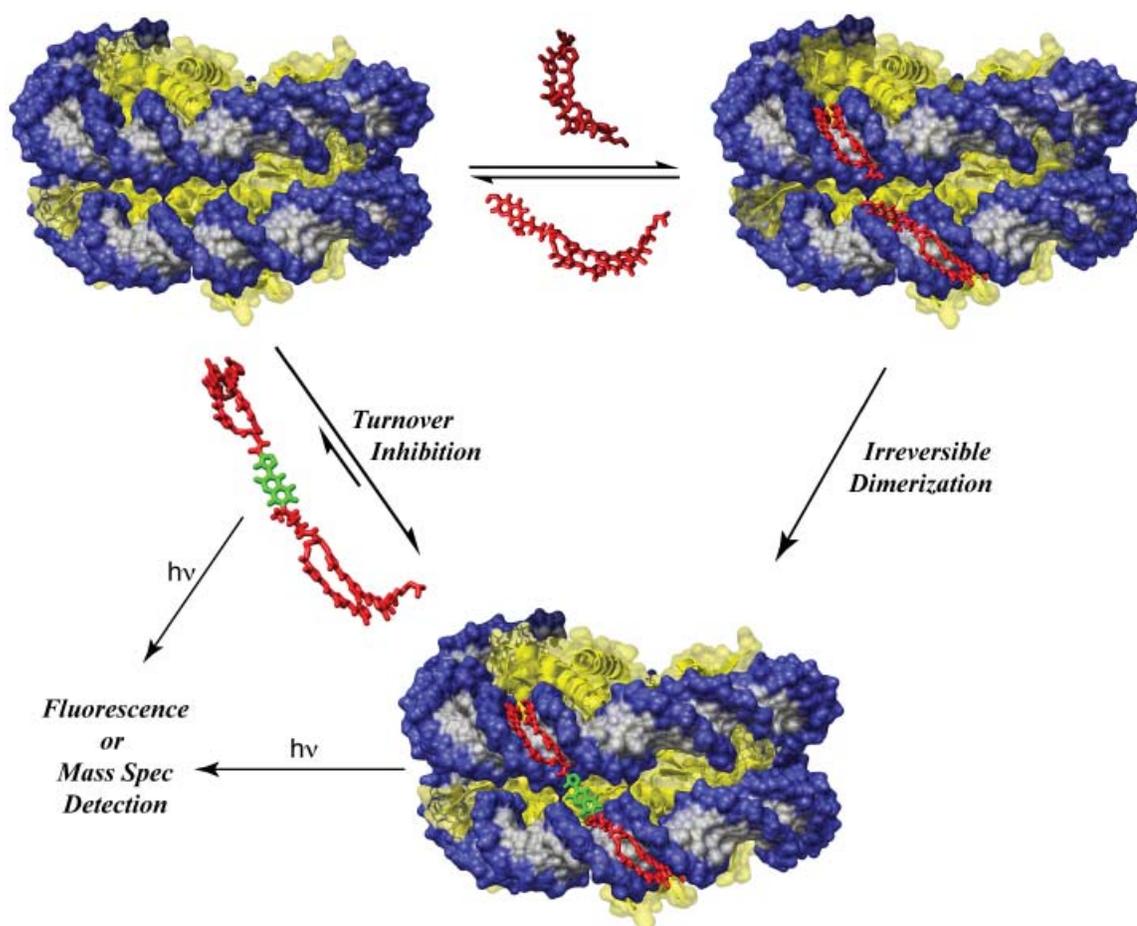
The ability to use double stranded DNA as a template for the ligation of two polyamides with bioorthogonal functionality has recently been demonstrated.<sup>1,2</sup> The 1,3-dipolar cycloaddition between an azide and alkyne was employed, resulting in a triazole linked tandem hairpin polyamide capable of targeting >10 base pairs of DNA with rate increases of 20,000-fold.<sup>1,2</sup> The ability to template the dimerization of two cell permeable molecules for the targeted downregulation of a gene represents a powerful strategy for the control of transcriptional regulated processes in biological systems. The Dervan laboratory has also shown that polyamides are capable of binding the nucleosome core particle (NCP) and that a polyamide clamp constructed from two 8-ring polyamides connected at the turns by a linker is capable of locking the DNA onto the histone octamer (Figure 9.1).<sup>3,4</sup> This NCP clamp was able to lock a complete turn of DNA onto the octamer significantly improving *in vitro* stability. Unfortunately, NCP clamps are very high molecular weight branched oligomers (MW > 2500) and often have poor cell permeability profiles.<sup>5,6</sup> The ability to template the dimerization of two hairpin polyamides across the supergroove between the two gyres of DNA on the NCP would allow for the introduction of two lower molecular weight polyamides potentially obviating cell permeability and size issues.

Incorporation of reporting strategies into the ligation event such as a profluorescent linker, where fluorescence would be activated by the dimerization process, would allow for the direct monitoring of reaction kinetics and cellular localization. The clamp dimerization strategy is illustrated in Figure 9.2 and would provide an important tool for the study of nucleosome core



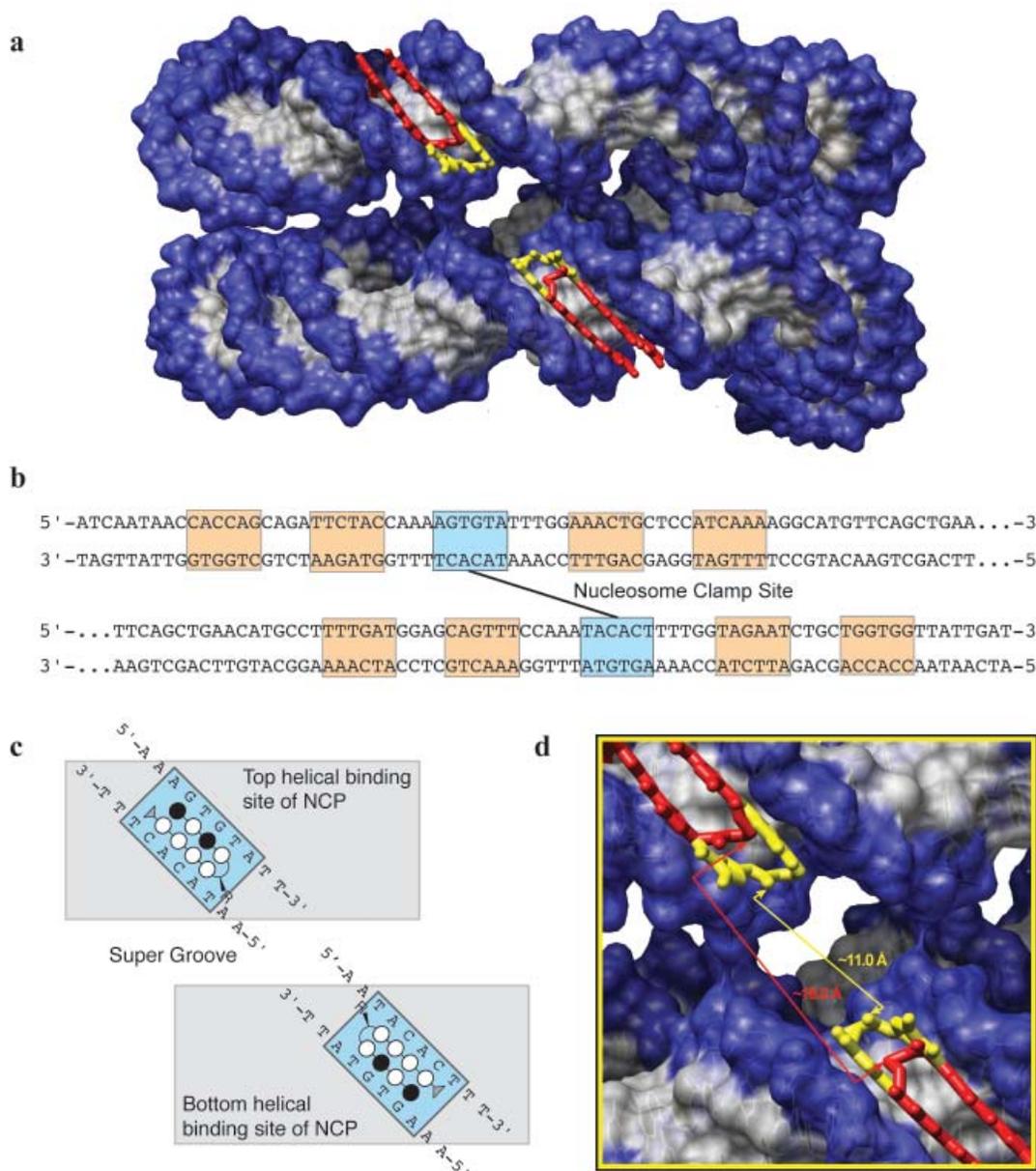
**Figure 9.1** High resolution crystal structure of an NCP bound polyamide clamp. Clamp is shown in red. DNA backbone shown in blue with minor and major grooves in gray. Side view (left) showing histone octamer core (orange, yellow, tan, and pink). Top view (right) with clamp bound in a minor supergroove on the proximal face. (PDB 1S32)

particle recognition and gene regulation in addition to serving as a macromolecular structural probe. The potential for gene regulation at the NCP level represents a unique physiologically relevant recognition platform distinct from that of linear DNA. In addition, NCP's tend to be sensitive to radiation damage and this along with dynamic DNA dissociation significantly reduce crystal diffraction quality.<sup>4</sup> Polyamide clamps provide a unique tool for structural biology due to their ability to increase the size, order, and resolution of NCP crystals, as evidenced by the 2 Å resolution structure in Figure 9.1, providing the potential for crystallization of larger physiologically relevant structures.<sup>4</sup> One could dream of the possibility of crystallizing structures such as multiply linked histone octamers or NCP's with bound transcription factors and clamps of this sort could be just the



**Figure 9.2** Illustration of the overall clamp dimerization strategy. Nucleosomal DNA shown without histone octamer present for clarity. Initial polyamide binding is a reversible process along with unwinding and winding of the DNA from the histone octamer. Dimerization of the two polyamides via a bioorthogonal reaction templated by the NCP is irreversible resulting in a large, sequence specific, high affinity clamp. The clamp can serve to stabilize the NCP for structural studies or potentially downregulate gene expression. Direct kinetic information and cellular localization data can be obtained via dimerization induced profluorescent linker activation.

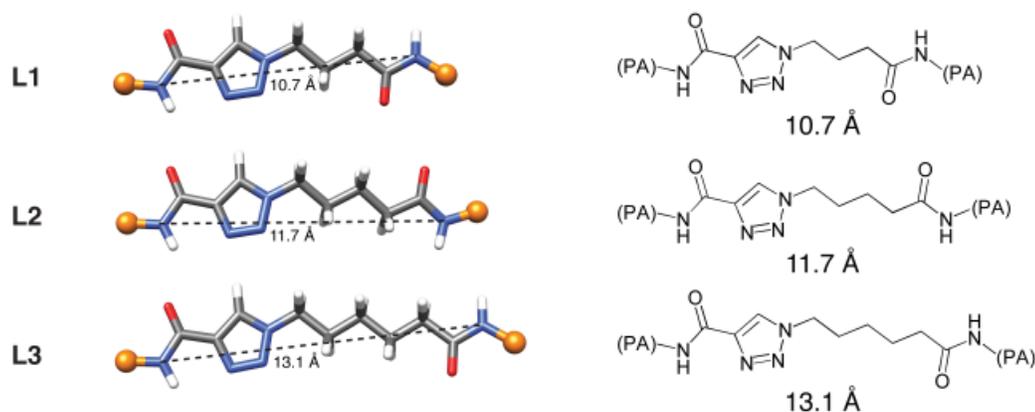
tool needed to gain structural insight into these complex biological architectures.



**Figure 9.3** a) Sequence of the 146 base pair fragment of  $\alpha$ -satellite DNA used for crystallographic studies with the nucleosome clamp. In those structures, the clamp was found to bind in the homodimeric “supergroove” highlighted in yellow. Each of the other four supergrooves on the NCP are highlighted in purple, green, blue and red. a) Highlight of the sequences to which the nucleosome clamp was bound (left). At right is shown the two base pair mutation (in red) introduced so that the supergroove becomes heterodimeric. Located above each highlighted site is the polyamide designed to target that site. b) Analysis of the linker dependence of NCP templated ligations. Crystal structure view of the supergroove. The predicted distance between the amines of the turn are shown in green for binding at the original site, and in red if the sites are moved one base pair back. c) Modeling of the linkers and the calculated turn-to-turn length.

## 9.2 Results and Discussion

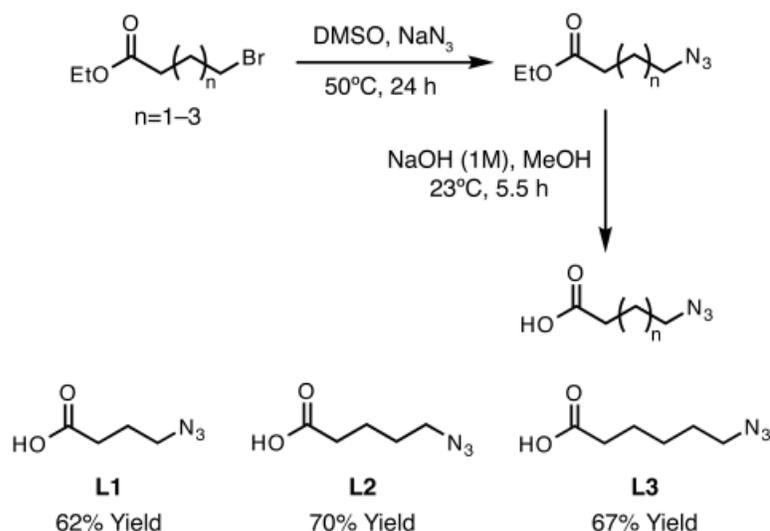
In order to ascertain whether the NCP could be used to template ligation reactions (Figure 9.3), a series of three azide containing polyamides and two alkyne containing polyamides were synthesized. Several linker lengths were used in order to examine the distance dependence of the reaction (Figure 9.4). Examination of the previous crystal structure data and computational modeling



**Figure 9.4** Analysis of the linker dependence of NCP templated ligations. Models and structure of the linkers with the calculated turn-to-turn length listed below each.

showed that only the azide containing the longest linker was expected to be capable of reacting. In addition, the previous study of DNA ligation had demonstrated that alkyne had reacted over 20,000 times faster than the alkyne containing an extra carbon unit and was expected to perform similarly in our experiments. The alkynes were purchased commercially and then coupled directly to the appropriate polyamides as described previously. The azides were synthesized using the scheme shown in Figure 9.5, 9.6, and 9.7.

The NCP was reconstituted as outlined in the supplemental information. A control sample containing only the 146 bp DNA and a second control containing

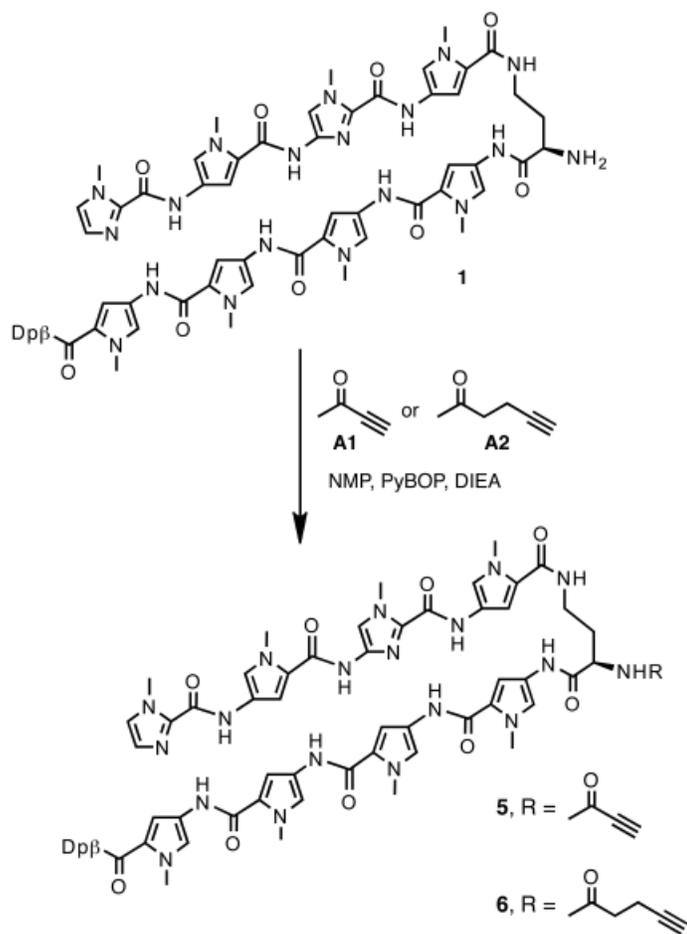


**Figure 9.5** Synthesis of alkyl azido linkers.<sup>6</sup>



the quantitation of ligation yield. As previously discussed, a fluorescent reporter strategy utilizing pro-fluorescent probes could provide an alternative to MALDI-MS analysis, with far less background noise.

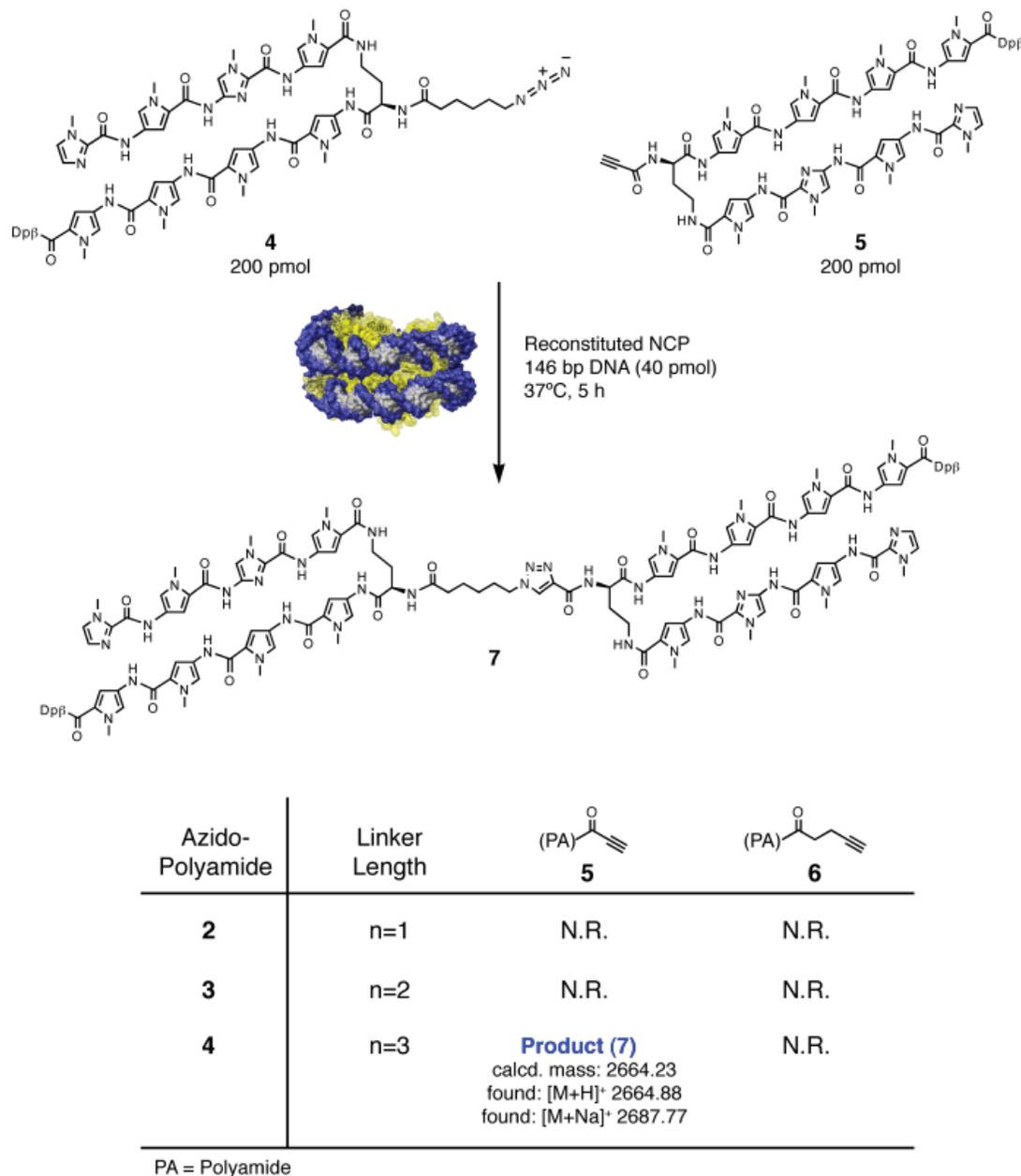
Coumarins have enjoyed widespread use as platforms for the discovery of fluorescent molecules (e.g. laser dyes). Reports of profluorescent coumarins, which upon chemical modification or reaction give rise to a fluorescent coumarin based molecules have been reported.<sup>7</sup> Azido coumarins have been shown to react with alkynes via 1,3-dipolar cycloadditions giving rise to highly fluorescent products due to a change in donor acceptor properties of appended functionality (Figure 9.10).<sup>7</sup> In addition, maleimide-functionalized coumarins are similarly non-fluorescent until conjugate addition with a thiol yields a fluorescent product.<sup>8</sup> These reactions could potentially be used in a bioorthogonal profluorescent ligation strategy for templating the dimerization of polyamides on the nucleosome core particle as illustrated in Figure 9.2. The ability of the linker to act as a fluorescent switch offers the unique possibility of monitoring reaction progress and cellular localization using highly



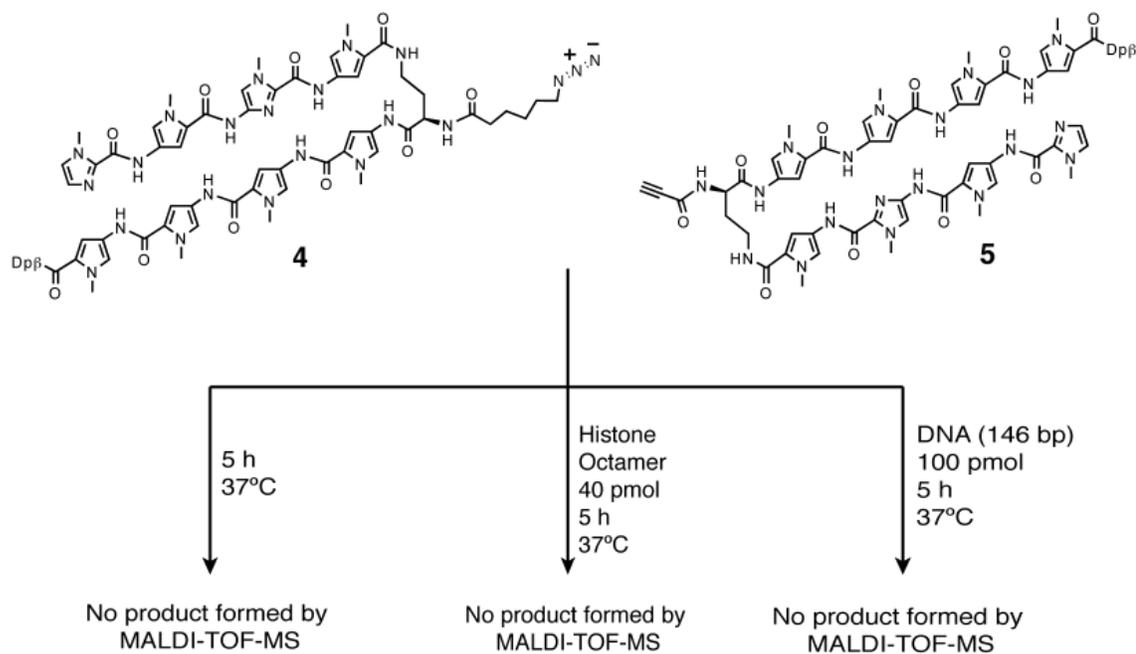
**Figure 9.7** Synthesis of alkynyl-polyamides.

sensitive fluorescent techniques. To establish linker requirements for the coumarins, molecular modeling was done using Spartan ES software package<sup>9</sup> and energy-minimized using an AM1 model, followed by ab initio calculations by means of the Hartree-Fock method using the 6-31G\* Pople basis set.<sup>10</sup> As shown in Figure 9.10, the fully compacted, shortest triazole-coumarin spans 13 Å, which is 2 Å longer than the 11.0 Å distance between the two 8-ring polyamide turn amines in the crystal structure (Figure 9.3). This posed a potential problem due to the rigidity of the of the coumarin system. When 6-ring hairpin polyamides were modeled in for the 8- ring polyamides on the

crystal structure, the distance between turn amines increased to 17.8 Å. Replacement of 8-ring polyamides for 6-ring polyamides was a potential solution to the problem, but not ideal since all structural data to date has been generated for 8-ring polyamides bound to the NCP. After surveying the literature for potential profluorescent replacements the aza-analogues of coumarins (carbostyrils) were discovered. Carbostyrils<sup>4,11</sup> exhibit photophysical properties similar to coumarins and can



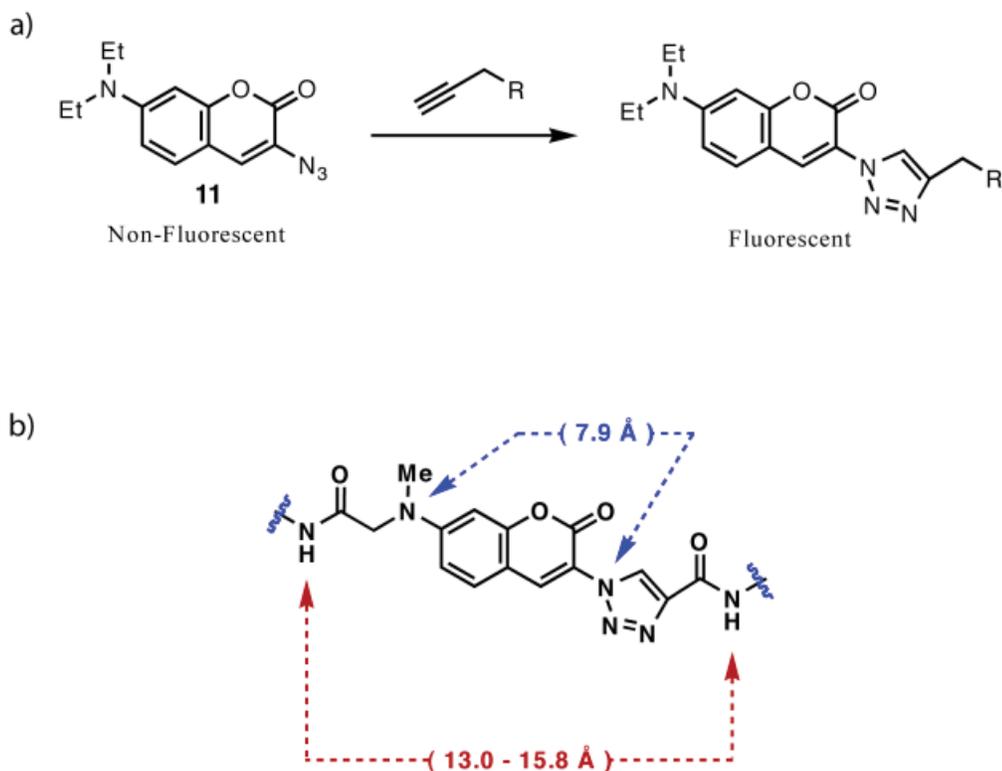
**Figure 9.8** DNA templated ligation on the NCP. The reaction of **4** and **5** with the NCP leads to dimer **7** which was observed by MALDI mass spectrometry. NCP templated ligation between different length azides and alkynes. The reaction product was only detected with polyamides **4** and **5**. N.R. = no reaction.



**Figure 9.9** Control reactions for the NCP templated ligation. The ligation reaction does not occur in the presence of just buffer, just the histone octamer, or just 146 bp DNA.

be electronically tuned by substitution with electron withdrawing and donating functionality.<sup>8,12</sup> Figure 9.11 shows the structure and linker distance comparison for carbostyrils versus coumarins. It appears that the carbostyril based linkers will be able to accommodate both 6- and 8-ring systems. In addition, synthetic ease, compact size, and a nitrogen handle for derivatization makes the carbostyrils an ideal candidate for use as a coumarin alternative. Preliminary efforts towards the synthesis of carbostyril (**17**) are presented in Figure 9.12.

Synthesis of carbostyril (**17**) (Figure 9.12) started with condensation of commercially available *N,N*-dimethyl-*m*-phenylenediamine (**8**) with dimethyl malonate at 200 °C to give 2,4-dihydroxy-7-(dimethylamino)quinoline (**14**) in 50% isolated yield after a single recrystallization. Chlorination of **14** under refluxing POCl<sub>3</sub> afforded 2,4-dichloro-7-(dimethylamino)quinoline (**15**) in 80% yield following recrystallization. Quinoline **15** was hydrolyzed under refluxing conditions in 6M HCl to afford 4-chloro-7-(dimethylamino)carbostyril (**16**) in 90% yield. Selective hydrolysis of the chlorine at position 2 is often explained by the observation that the 2 position is more reactive toward nucleophilic substitution than the 4 position. After comparing the proton NMR of compound **16** to that reported in the literature, an apparent discrepancy was realized. Since there was a possibility of displacement of chlorine at two different positions and NMR would not unambiguously resolve the regioselectivity issue, the compound was recrystallized from hot DMF

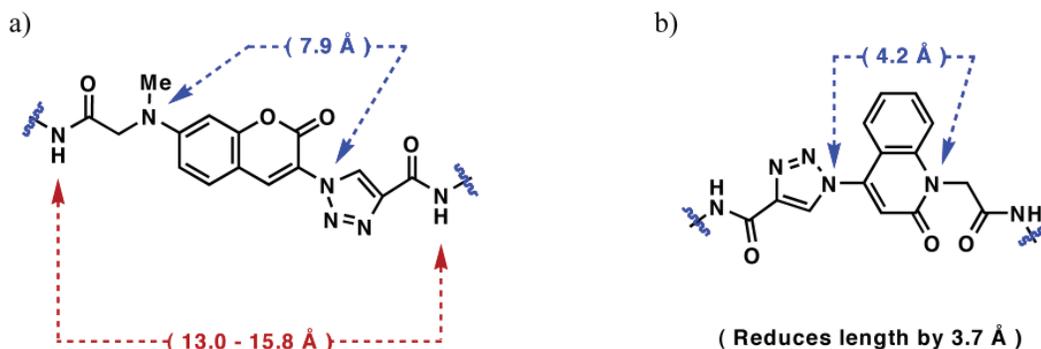


**Figure 9.10** a) Reaction of a profluorescent azido coumarin giving rise to a fluorescent triazole product. b) Linker distances for incorporation of the profluorescent coumarin.

and the crystal structure solved. The crystal structure and unit cell are shown in Figure 9.12. After unambiguously proving the identity of compound **16** it was subjected to azidification. Compound **16** was treated with excess sodium azide in  $d_6$ -DMSO and the reaction monitored by proton NMR. After heating for exactly 10.5 h at 120 °C complete conversion to azide **17** was achieved by NMR. Compound **17** can be N-alkylated to provide a functional handle for conjugation to small molecules and macromolecules. Initial investigations into the photophysical properties of compound **17** are shown in Figure XX and demonstrate the ability of the probe to fluoresce upon cycloaddition with an alkyne substrate.

### 9.3 Conclusion

These studies have demonstrated the feasibility of the NCP templated ligation approach for the self assembly of polyamide dimers. Additionally, the development of a pro-fluorescent azido carbostyryl provides a new tool for monitoring ligation reactions using fluorescence. This strategy offers an exciting opportunity for modifying gene expression in cells by the targeted self-assembly of polyamides on NCP's.



**Figure 9.11** a) Coumarin linker length presented for comparison against carboxytril. b) Carboxytril linker distance shows a decrease in length of ~3.7 angstroms.

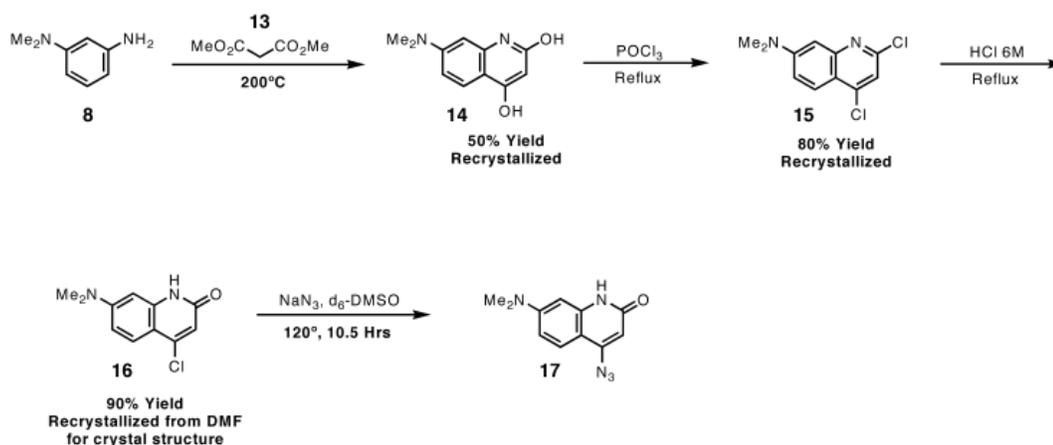
## 9.4 Experimental

### 9.4.1 Materials and General Methods

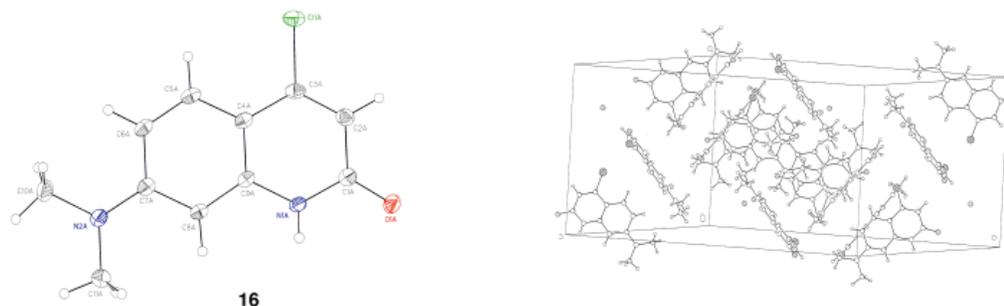
Dicyclohexylcarbodiimide (DCC), 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), *N*-Hydroxybenzotriazole (HOBt), Fmoc-Dab(Boc)-OH and Boc- $\beta$ -Ala-Pam resin were purchased from Peptides International. Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) was purchased from Novabiochem. Fluoro-*N,N,N,N'*-tetramethylformamidinium hexafluorophosphate (TFFH) was purchased from Advanced ChemTech. O-(7-Azabenzotriazol-1-yl)-*N,N,N,N'*-tetramethyluronium hexafluorophosphate (HATU), 4-(dimethylamino)-pyridine (DMAP), *N*-hydroxysuccinimide (NHS), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidinone (NMP), *N,N*-dimethylpropylamine (Dp), *N,N*-diisopropylethylamine (DIEA), ethylene diamine, piperidine, and other miscellaneous chemicals were purchased from Aldrich and used without further purification. All other solvents were purchased from EM Sciences and were reagent grade. Trifluoroacetic acid (TFA) was purchased from Halocarbon.

$^1\text{H}$  NMR spectra were recorded using a 300MHz General Electric-QE NMR spectrometer.  $\text{CDCl}_3$  was obtained from Cambridge Isotope Laboratories. UV spectra were recorded in water using a Beckman Coulter DU 7400 Spectrophotometer. Matrix-assisted LASER desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) was performed using an Applied Biosystems Voyager DE Pro Spectrometer. Electrospray ionization (ESI) mass spectrometry was performed by the Protein and Peptide Microanalytical Facility at the California Institute of Technology. Analytical High-Pressure Liquid Chromatography (HPLC) was performed with a Beckman Gold system using a Varian Microsorb-MV 100 C18 column (5 $\mu\text{m}$  particle size, 250 x 4.6mm). Preparative HPLC was performed using a Beckman Gold system with either a Waters Bondapak C18 column (15-

a)



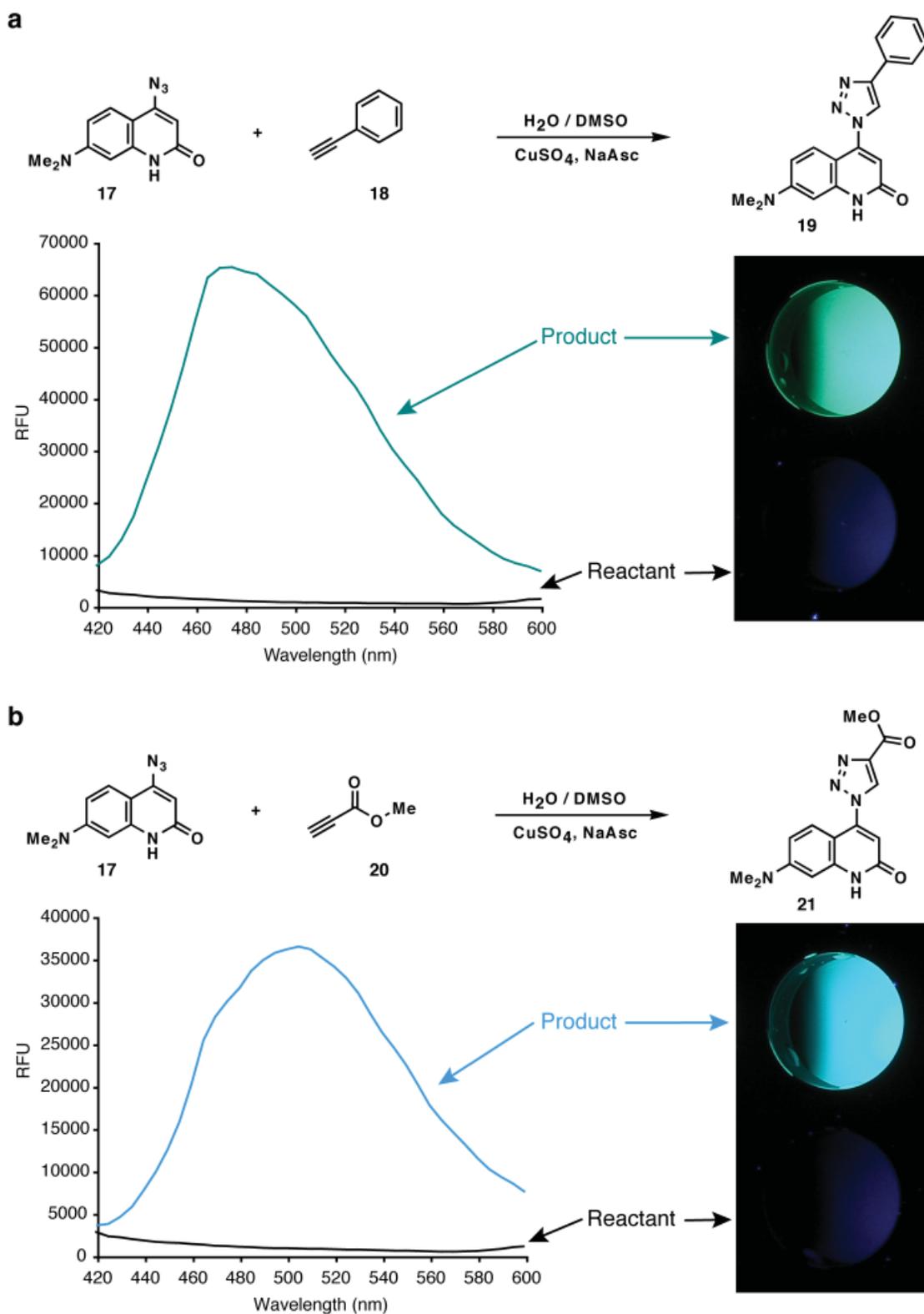
b)



**Figure 9.12** a) Progress toward the synthesis of profluorescent azide **18**. b) X-ray crystal structure and unit cell for structural proof of compound **16**.

20  $\mu\text{m}$  particle size, 25 x 100mm) or a Phenomenex Gemini C18 column ( 5  $\mu\text{m}$  particle size, 250 x 21.2 mm). For both HPLC systems Solvent A was 0.1% (v/v) aqueous TFA and solvent B was acetonitrile. Analytical HPLC was done using a gradient of 1.85%/min of Solvent B starting from 0% over 35 min with a flowrate of 1.5mL/min. Preparative HPLC was typically done using a gradient of 1%/min of Solvent B for 20 min followed by a gradient of 0.3%/min for an additional 100min at a flowrate of 8mL/min. Radioactive gels were imaged using a Molecular Dynamics 400S PhosphorImager.

Restriction endonucleases, deoxyribonucleotide triphosphates, DNase I, Polynucleotide kinase (PNK), and glycogen were purchased from Roche. [ $\alpha$ - $^{32}\text{P}$ ]-Thymidine-5'-triphosphate ( $\geq 3000$  Ci/mmol) and [ $\alpha$ - $^{32}\text{P}$ ]-Deoxyadenosine-5'-triphosphate ( $\geq 6000$  Ci/mmol) were purchased from Perkin Elmer. [ $\gamma$ - $^{32}\text{P}$ ]-Adenosine-5'-triphosphate ( $\geq 7000$  Ci/mmol) was purchased from MP Biomedicals. Water was purified from a Millipore Mill-Q purification system for general use.



**Figure 9.13** Photophysical properties of compound **17** before and after cycloaddition with alkynes **18** and **20**.

Ultrapure RNase/DNase free water from USB was used for biological work. All buffer reagents used were molecular biology grade. Buffers were sterilized using a Nalgene 0.2 $\mu$ m cellulose filtration device.

#### 9.4.2 Plasmids

pJDC1 and pJDC2 were constructed using 80mer oligonucleotides purchased from Integrated DNA Technologies. pUC19 plasmid was purchased from Sigma. JM109 Competent Cells (>108 cfu/ $\mu$ g) were purchased from Promega. A Rapid DNA ligation kit was purchased from Roche. Purification was done using a Promega Wizard Plus Midipreps DNA purification system. Sequence analysis was performed by the Sequence Analysis Facility at the Caltech.

#### 9.4.3 Polyamide Synthesis

Polyamide synthesis was performed as previously reported.<sup>13</sup> All polyamides were synthesized using Boc- $\beta$ -Ala-PAM resin (~0.59 meq/g). The resin was initially swelled in DMF for 5 min in a glass reaction vessel fitted with a glass filter and stopcock. The vessel was drained and the resin washed twice with DCM. Deprotection with 80% TFA:DCM was performed for 20 min while the resin was shaken. Following deprotection, the resin was washed 2 x DCM, 1 x 4:1 DMF:DIEA, and 1 x DMF. Coupling of the Boc-Py-OBt pre-activated ester was performed using 1.8 eq of monomer in ~1 mL of NMP. Coupling of Boc-Im-OH, Boc-PyIm-OH, and other acids was done by first preactivating 1.8 eq of the monomer with 1.7 eq of HBTU, 5.4 eq of DIEA in ~2mL of NMP. The activation mixture was shaken for 20min before being filtered and added to the resin. Couplings were allowed to proceed for 2h except in the case of Im-OH which was allowed to react overnight. After each coupling step the resin was washed 2 x DMF and then 2 x DCM before the next deprotection step. Polyamides were cleaved from resin using 1.5 mL of Dp for 200 mg of resin at 55°C for eight h. Crude products were purified by preparative HPLC.

2: MALDI-TOF-MS calculated [M+H]<sup>+</sup>: 1348.63, observed 1348.54

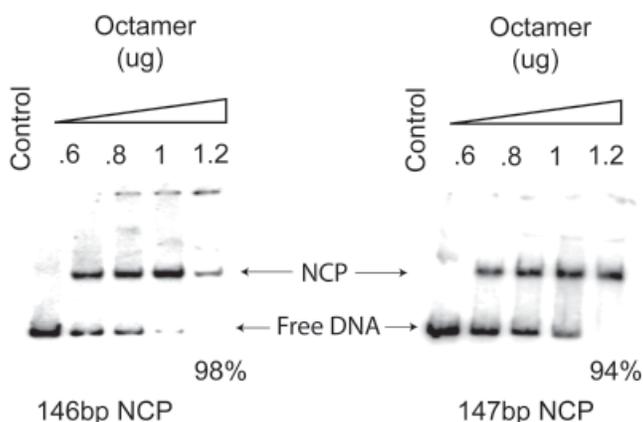
3: MALDI-TOF-MS calculated [M+H]<sup>+</sup>: 1362.65, observed 1362.61

4: MALDI-TOF-MS calculated [M+H]<sup>+</sup>: 1376.66, observed 1376.73

5: MALDI-TOF-MS calculated [M+H]<sup>+</sup>: 1289.58, observed 1289.67

6: MALDI-TOF-MS calculated [M+H]<sup>+</sup>: 1317.61, observed 1317.6

7: MALDI-TOF-MS calculated [M+H]<sup>+</sup>: 2664.23, observed 2664.88



**Figure 9.14** Representative gels of the NCP reconstitution. The amount of histone octamer is titrated into a constant amount of the 146 or 147 bp radiolabelled DNA. Non-denaturing polyacrylamide gels are used to determine the yield from reconstitution.

#### 9.4.4 Reconstitution of NCP

The NCP was reconstituted following established protocols. Briefly, 10  $\mu\text{g}$  of the 146 bp DNA was 5' radiolabelled as described above using Polynucleotide Kinase and purified using a Chroma Spin STE 10 column from BD Biosciences. Histone octamer which was obtained from the Luger lab was diluted to 1  $\mu\text{g}$  /  $\mu\text{L}$  using 2 M NaCl TE Buffer. 5  $\mu\text{L}$  of the DNA was added to 5  $\mu\text{L}$  of 4M NaCl, 10 mM Tris, 1 mM EDTA solution to make a sample of 10  $\mu\text{L}$  of DNA in 2 M NaCl TE buffer. Five samples were

made up and 0, 0.6, 0.8, 1.0, or 1.2  $\mu\text{L}$  of octamer (1  $\mu\text{g}$  /  $\mu\text{L}$ ) was added to each sample. The total volume was brought to 2  $\mu\text{L}$  using 2M NaCl TE buffer. The samples were incubated for 1 hr before the addition of 12  $\mu\text{L}$ , 6  $\mu\text{L}$ , 6  $\mu\text{L}$ , and 84  $\mu\text{L}$  of dilution buffer (TE) with each addition followed by a 1 h incubation. The sample was heated to 37°C for 2 h before a final addition of 120  $\mu\text{L}$  of dilution buffer to give a final NaCl concentration of .1 M. Reconstituted NCP samples were stored at 4°C. Non-denaturing PAGE was done on a 6% polyacrylamide TBE gel. 1  $\mu\text{L}$  of each sample was diluted to 10  $\mu\text{L}$  with 10 mM Tris, 20 mM NaCl, 10% glycerol, and .1% Igapal. The gel was run at 150 V for 20 min, dried and imaged.

#### 9.4.5 NCP Templated Ligation Reactions

The NCP was reconstituted as described above with the sole difference that non-radiolabelled DNA was used. For each ligation reaction 200 pmol of each polyamide was lyophilized into an eppendorf tube and 40 pmol of the reconstituted NCP was added. Samples containing only the 146 bp DNA and or the histone octamer were prepared as negative controls. The reaction was incubated for 5 h at 37°C before MALDI-TOF analysis.

### 9.5 Notes and References

1. Poulin-Kerstien, A. T., and Dervan, P. B. DNA-templated dimerization of hairpin polyamides. *J. Am. Chem. Soc.* **2003**, *125*, 15811-15821.

2. Poulin-Kerstien, A. T. *DNA-Templated Dimerizations of Minor Groove--Binding Polyamides* California Institute of Technology: Pasadena, CA, **2005**
3. Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* **2003**, *326*, 371-380.
4. Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Molecular recognition of the nucleosomal "super groove". *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6864-6869.
5. Best, T. P., Edelson, B. S., Nickols, N. G., and Dervan, P. B. Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12063-12068.
6. Khoukhi, N., Vaultier, M., and Carrie, R. Synthesis and reactivity of methyl-azido butyrates and ethyl-azido valerates and of the corresponding acid chlorides as useful reagents for the aminoalkylation *Tetrahedron.* **1987**, *43*, 1811-1822.
7. Sivakumar, K., Xie, F., Cash, B. M., Long, S., Barnhill, H. N., and Wang, Q. A Fluorogenic 1, 3-Dipolar Cycloaddition Reaction of 3-Azidocoumarins and Acetylenes *Org. Lett.* **2004**, *6*, 4603-4606.
8. Chaurasia, C. S., and Kauffman, J. M. Synthesis and fluorescent properties of a new photostable thiol reagent "BACM." *J. Heterocyc. Chem.* **1990**, *27*, 727-733.
9. *Spartan ES Wavefunction*, Inc: Irvine, CA, **2005**.
10. (a) Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., Defrees, D. J., and Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The. Journal. of. Chemical. Physics.* **1982**, *77*, 3654-3665. (b) Hariharan, P. C., and Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretical. Chemistry. Accounts: Theory, Computation, and Modeling. (Theoretica. Chimica. Acta).* **1973**, *28*, 213-222.
11. Edayathumangalam, R. S., Weyermann, P., Dervan, P. B., Gottesfeld, J. M., and Luger, K. Nucleosomes in solution exist as a mixture of twist-defect states. *J. Mol. Biol.* **2005**, *345*, 103-114.
12. Nasr, M., Drach, J. C., Smith, S. H., Shipman Jr, C., and Burckhalter, J. H. 7-Aminoquinolines. A novel class of agents active against herpesviruses. *J. Med. Chem.* **1988**, *31*, 1347-1351.
13. Baird, E. E., and Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

## **Appendix A: Next Generation Hairpin Polyamides with (R)-3,4-**

### **Diaminobutyric Acid Turn Unit**

*The text of this chapter was taken in part from a manuscript coauthored with Christian Dose, Michelle E. Farkas, and Peter B. Dervan\* (Caltech)*

(Dose, C., Farkas, M.E., Chenoweth, D.M., and Peter B. Dervan *J. Am. Chem. Soc.*, **2008**, *130*, 6859-6866.)

**Abstract**

The characterization of a new class of pyrrole–imidazole hairpin polyamides with  $\beta$ -amino- $\gamma$ -turn units for recognition of the DNA minor groove is reported. A library of eight hairpins containing (R)- and (S)-3,4-diaminobutyric acid ( $\beta$ -amino- $\gamma$ -turn) has been synthesized, and the impact of the molecules on DNA-duplex stabilization was studied for comparison with the parent  $\gamma$ -aminobutyric acid ( $\gamma$ -turn) and standard (R)-2,4-diaminobutyric acid ( $\alpha$ -amino- $\gamma$ -turn)-linked eight-ring polyamides. For some, but not all, sequence compositions, melting temperature analyses have revealed that both enantiomeric forms of the  $\beta$ -amino- $\gamma$ -turn increase the DNA-binding affinity of polyamides relative to the (R)- $\alpha$ -amino- $\gamma$ -turn. The (R)- $\beta$ -amine residue may be an attractive alternative for constructing hairpin polyamide conjugates. Biological assays have shown that (R)- $\beta$ -amino- $\gamma$ -turn hairpins are able to inhibit androgen receptor-mediated gene expression in cell culture similar to hairpins bearing the standard (R)- $\alpha$ -amino- $\gamma$ -turn, from which we infer they are cell-permeable.

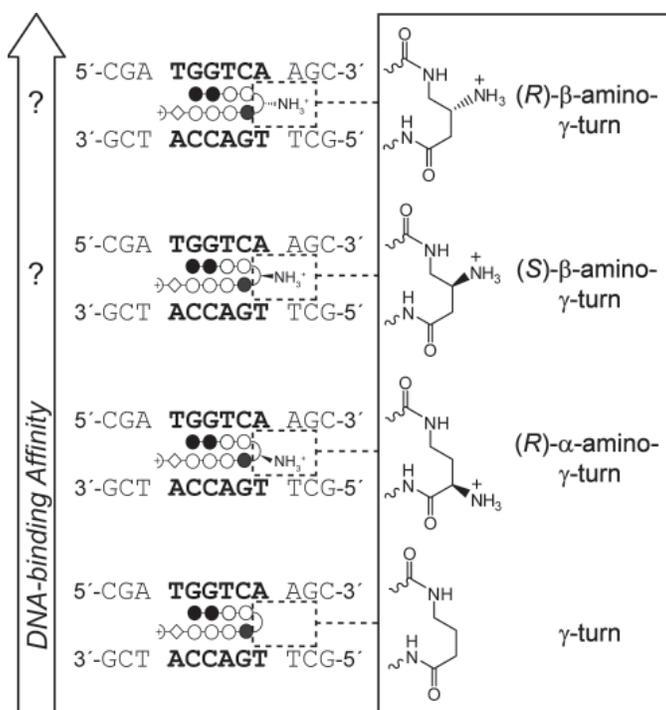
## A.1 Introduction

The ability to modulate the expression of eukaryotic gene networks by small molecules is a challenge in the field of chemical biology. Hairpin pyrrole-imidazole polyamides are a class of programmable small molecules that bind to the minor groove of DNA with affinities similar to transcription factors and have been shown to inhibit gene expression in living cells by interfering with transcription factor/DNA interfaces.<sup>1</sup> The DNA sequence specificity of polyamides arise from interactions of pairs of the aromatic amino acids *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methylhydroxypyrrole (Hp) with the edges of the Watson-Crick base pairs.<sup>2</sup> The generality of the polyamide pairing rules has been demonstrated by numerous studies<sup>3</sup> and applications of polyamide conjugates include DNA alkylations,<sup>4</sup> DNA-templated ligations,<sup>5</sup> sequence-specific DNA intercalators,<sup>6</sup> fluorescent DNA paints,<sup>7</sup> DNA nanoarchitectures,<sup>8</sup> and transcription factor mimics.<sup>9</sup> Efforts have been made to improve the DNA-binding properties of hairpin polyamides with modified turn units.<sup>10</sup> Substitution of  $\gamma$ -aminobutyric acid ( $\gamma$ -turn) by (R)-2,4-diaminobutyric acid ( $\alpha$ -amino- $\gamma$ -turn) increases the DNA-binding affinity by  $\sim 15$ -fold.<sup>10b,11</sup> In contrast, hairpins containing the opposite enantiomer, (S)- $\alpha$ -amino- $\gamma$ -turn, result in diminished binding affinities. This decrease is most likely caused by an unfavorable steric clash of the amine residue with the DNA minor groove.<sup>10b</sup> Sugiyama and co-workers have introduced polyamides containing the  $\alpha$ -hydroxy- $\gamma$ -turn.<sup>10c</sup> These hairpins provide discrimination for A•T/T•A base pairs at the turn position, although a  $\sim 70$ -fold reduced DNA-binding affinity relative to analogue (R)- $\alpha$ -amino- $\gamma$ -turn-linked polyamides has been observed.

Here we introduce a new class of hairpin polyamides which are linked by 3,4-diaminobutyric acid which results in a  $\beta$ -amine residue at the turn unit ( $\beta$ -amino- $\gamma$ -turn) (Figure A.1). DNA-binding affinities of four different eight-ring polyamide core sequences (with incrementally increasing Im/Py pairs) have been investigated and were compared to analogue hairpins bearing the parent  $\gamma$ -turn and the standard (R)- $\alpha$ -amino- $\gamma$ -turn. We show that, for certain series of hairpin polyamides, both enantiomers of the  $\beta$ -amino- $\gamma$ -turn are able to increase the relative DNA-binding affinity. However, this is sequence context dependent. Biological assays revealed that hairpin polyamides bearing the (R)- $\beta$ -amino- $\gamma$ -turn are able to inhibit specific gene expression in cell culture, which is taken as evidence for cell permeability.

## A.2 Results and Discussion

### A.2.1 Thermal stabilization of DNA duplexes by hairpin polyamides

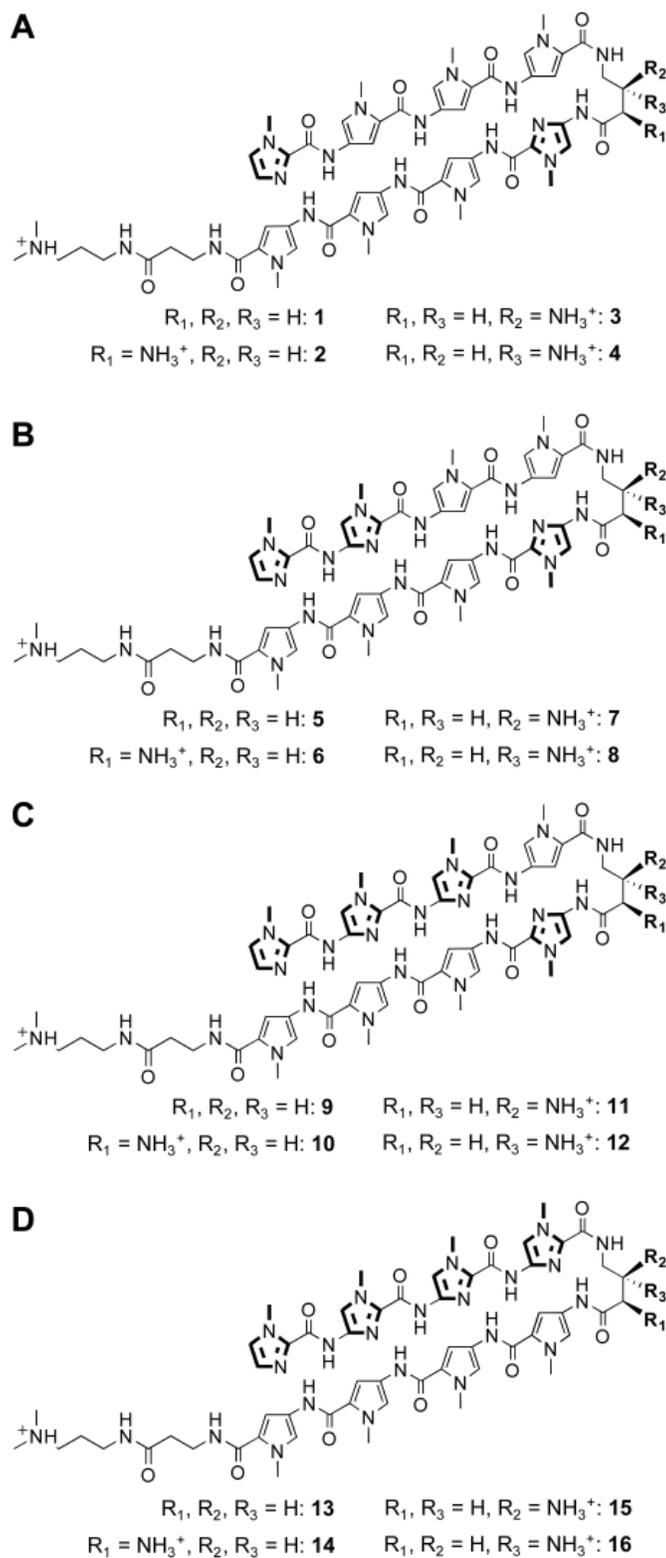


**Figure A.1** Schematic representation of hairpin polyamides with increased DNA-binding affinity caused by different  $\gamma$ -turn units. Hairpin polyamides targeted to DNA sequence 5'-TGGTCA-3' are shown as ball-and-stick models. Ball-and-stick representation legend: black and white circles represent *N*-methylimidazole and *N*-methylpyrrole units, respectively, half-circles represent  $\gamma$ -aminobutyric acid, white diamonds represent  $\beta$ -alanine units, and half-circles containing a cross represent 3-(dimethylamino)-1-propylamine (Dp) as tail.

Hairpin polyamides **1-16** were synthesized with different Im/Py and Py/Py compositions targeted to the four DNA sequences with increasing G/C content 5'-TGTTC A-3', 5'-TGGTCA-3', 5'-TGGGCA-3', and 5'-TGGGGA-3' (Figure A.2). The energetics of DNA-binding properties of polyamides are typically characterized by quantitative DNase I footprint titrations.<sup>12</sup> These measurements provide precise information regarding the affinity and specificity of DNA/polyamide complexes. Unfortunately,

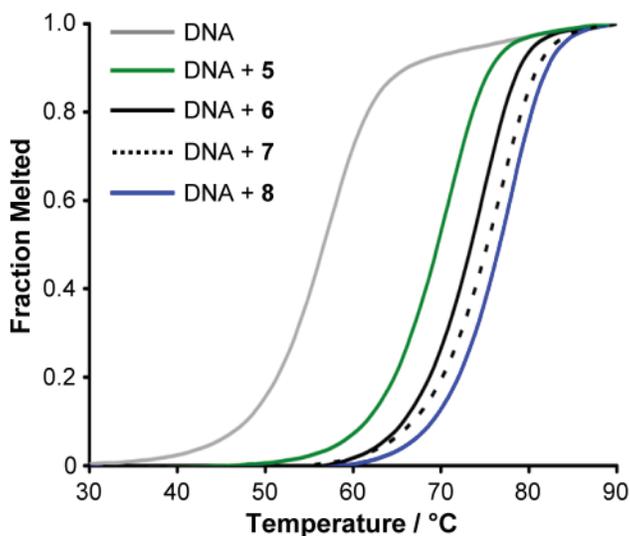
quantitative footprinting experiments revealed similar equilibrium association constants ( $K_a$  values  $\sim 2 \times 10^{10} \text{ M}^{-1}$ ) for hairpins **1-8**, reaching an upper limit of the standard procedure<sup>12,13</sup> (see Section A.6 Supplemental Information). Prior results have shown that the increase in melting temperature ( $\Delta T_m$ ) of DNA

duplexes bound by hairpin polyamides correlates with DNA-binding affinity and can be utilized to detect single base pair mismatched DNA/polyamide complexes.<sup>14</sup> Accordingly, we have used melting temperature analysis for dissecting differences in DNA affinities of hairpin polyamides. Spectroscopic analyses were performed on 12mer DNA duplexes containing the appropriate match sequence in the absence and presence of polyamides in order to derive the desired  $\Delta T_m$  values (Figure A.3). Table A.1 shows that all hairpins provided an increase in melting temperature, relative to the individual DNA duplexes, confirming the formation of DNA/polyamide complexes. As expected, spectroscopic analysis with (R)- $\alpha$ -amino- $\gamma$ -turn hairpins revealed stronger stabilizations than the parent  $\gamma$ -turn analogues; for example, achiral polyamide **1** targeted to DNA sequence 5'-TGTTC A-3' resulted in a  $\Delta T_m$  value of 15.9°C, while chiral hairpin (R)- $\alpha$ -**2** led to a 3.6°C higher



**Figure A.2** Chemical structures for hairpins **1-16** targeted to DNA sequences: (A) 5'-TGTTC A-3', (B) 5'-TGTGTC A-3', (C) 5'-TGGGCA-3', and (D) 5'-TGGGGA-3'.

melting temperature ( $\Delta T_m = 19.5^\circ\text{C}$ ). Remarkably, melting temperature analyses in the presence of  $\beta$ -amino- $\gamma$ -turn hairpins (S)- $\beta$ -**3** ( $\Delta T_m = 20.9^\circ\text{C}$ ) and (R)- $\beta$ -**4** ( $\Delta T_m = 22.2^\circ\text{C}$ ) revealed higher  $\Delta T_m$  values compared to those for the  $\alpha$ -series hairpin (R)- $\alpha$ -**2** ( $\Delta T_m = 19.5^\circ\text{C}$ ). The same trend was observed for hairpins **5-8** targeted to DNA sequence 5'-TGGTCA-3' (Table A.1, Figure A.2). First, it is noteworthy that both the (R)- and (S)- $\beta$ -amino- $\gamma$ -turn generated higher melting temperatures than the standard (R)- $\alpha$ -amino- $\gamma$ -turn. Second, the enhancement (relative to achiral hairpins) observed for the (R)- $\beta$ -series is almost twice that of the (R)- $\alpha$ -series targeted to DNA sequences 5'-TGTTC A-3' and 5'-TGGTCA-3'; for example, polyamide (R)- $\beta$ -**8** provided a  $\Delta\Delta T_m$  value of  $6.9^\circ\text{C}$ , while the  $\alpha$ -series (R)- $\alpha$ -**6** led to a  $\Delta\Delta T_m$  value of  $3.5^\circ\text{C}$  relative to achiral hairpin **5**. Interestingly, by further increasing the amounts of Im/Py pairs in the polyamides, significantly less DNA duplex stabilizations have been observed. For example, achiral polyamide **9** and chiral hairpin (R)- $\alpha$ -**10** targeted to DNA sequence 5'-TGGGCA-3' yielded  $\Delta T_m$  values of  $8.6$  and  $13.2^\circ\text{C}$ , while the  $\beta$ -series



**Figure A.3** Normalized UV denaturation profiles of 12mer DNA duplex 5'-CGATGGTCAAGC-3'/5'-GCTTGACCATCG-3' in the absence and presence of hairpin polyamides **5-8**.

(S)- **$\beta$ -11** and (R)- **$\beta$ -12** led to  $\Delta T_m$  values of 13.3 and 13.6°C, respectively (Table A.1). Even lower melting temperatures were observed for hairpins **13-16** designed to bind DNA sequence 5'-TGGGGA-3'. Both  $\beta$ -amino- $\gamma$ -turns, as in (S)- **$\beta$ -15** ( $\Delta T_m = 6.7^\circ\text{C}$ ) and (R)- **$\beta$ -16** ( $\Delta T_m = 6.8^\circ\text{C}$ ), resulted in significantly lower  $\Delta T_m$  values than the  $\alpha$ -series analogue (R)- **$\alpha$ -14** ( $\Delta T_m = 9.1^\circ\text{C}$ ). These results imply that the impact of polyamide turn units on DNA-duplex stabilization is sequence context dependent.

The general increase in DNA-binding affinity for polyamides containing the (R)- $\alpha$ -substituted  $\gamma$ -turn, relative to achiral hairpins, is most likely caused by a superposition of favorable noncovalent interactions of the positively charged substituent and conformational preferences of the turn unit.<sup>10b</sup> The (R)- $\alpha$ -amino- $\gamma$ -turn can exist in two different conformations, one orienting the  $\alpha$ -ammonium in a pseudoequatorial position (Figure A.4A), which directs the substituent toward the wall of the minor groove with the potential of steric interactions. The alternate conformation places the  $\alpha$ -amine residue in a pseudoaxial position out of the minor groove, orienting the  $\beta$ -methylene to the floor of the double helix (Figure A.4B). Modeling of the (S)- $\beta$ -amino- $\gamma$ -turn conformations suggests that the  $\alpha$ -ammonium in a pseudoaxial position is directed out of the minor groove (Figure A.4C) relieving the potential steric interactions with the wall in comparison to the  $\alpha$ -series. In contrast, the (S)- $\beta$ -amine in a pseudoequatorial orientation is following the curvature of the minor groove (Figure A.4D). The possibility for favorable noncovalent interactions should exist in both conformations without the detriment of steric interactions. As shown in Figure A.4E, the pseudoequatorial  $\beta$ -amine residue of the (R)- $\beta$ -amino- $\gamma$ -turn is well accommodated in the DNA minor groove, while the pseudoaxial position should result in a steric clash of the substituent with the groove floor (Figure A.4F). Previous results have shown that polyamides constructed with several continuous Im/Py pairs are overcurved with respect to the DNA minor groove, significantly influencing the DNA binding affinity and sequence specificity.<sup>15</sup> We assume that this curvature affects the alignment of the turn units in the DNA minor groove.

**Table A.1** Melting temperatures of DNA/polyamide complexes for A•T and T•A base pairs at the turn position of hairpin polyamides.<sup>a</sup>

Polyamides	A•T		T•A		
	5'-CGA <b>TGTTCA</b> AGC-3'	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$	5'-CGA <b>TGTTCT</b> AGC-3'	$T_m / ^\circ\text{C}$
—	54.0 (±0.2)	—	n.d.		
 (1)	69.9 (±0.3)	15.9	n.d.		
 (2)	73.5 (±0.2)	19.5	n.d.		
 (3)	74.9 (±0.2)	20.9	n.d.		
 (4)	76.2 (±0.2)	22.2	n.d.		
—	57.2 (±0.1)	—	55.8 (±0.1)	—	
 (5)	70.6 (±0.2)	13.4	69.0 (±0.3)	13.2	
 (6)	74.1 (±0.3)	16.9	72.9 (±0.2)	17.1	
 (7)	76.1 (±0.2)	18.9	73.2 (±0.1)	17.4	
 (8)	77.5 (±0.3)	20.3	74.2 (±0.1)	18.4	
—	60.2 (±0.2)	—	59.8 (±0.3)	—	
 (9)	68.8 (±0.2)	8.6	67.4 (±0.3)	7.6	
 (10)	73.4 (±0.2)	13.2	72.0 (±0.1)	12.2	
 (11)	73.5 (±0.1)	13.3	70.5 (±0.3)	10.7	
 (12)	73.8 (±0.1)	13.6	71.3 (±0.3)	11.5	
—	57.5 (±0.1)	—	57.9 (±0.1)	—	
 (13)	60.9 (±0.1)	3.4	61.4 (±0.3)	3.5	
 (14)	66.6 (±0.1)	9.1	67.0 (±0.1)	9.1	
 (15)	64.2 (±0.1)	6.7	64.2 (±0.1)	6.3	
 (16)	64.3 (±0.3)	6.8	64.4 (±0.3)	6.5	

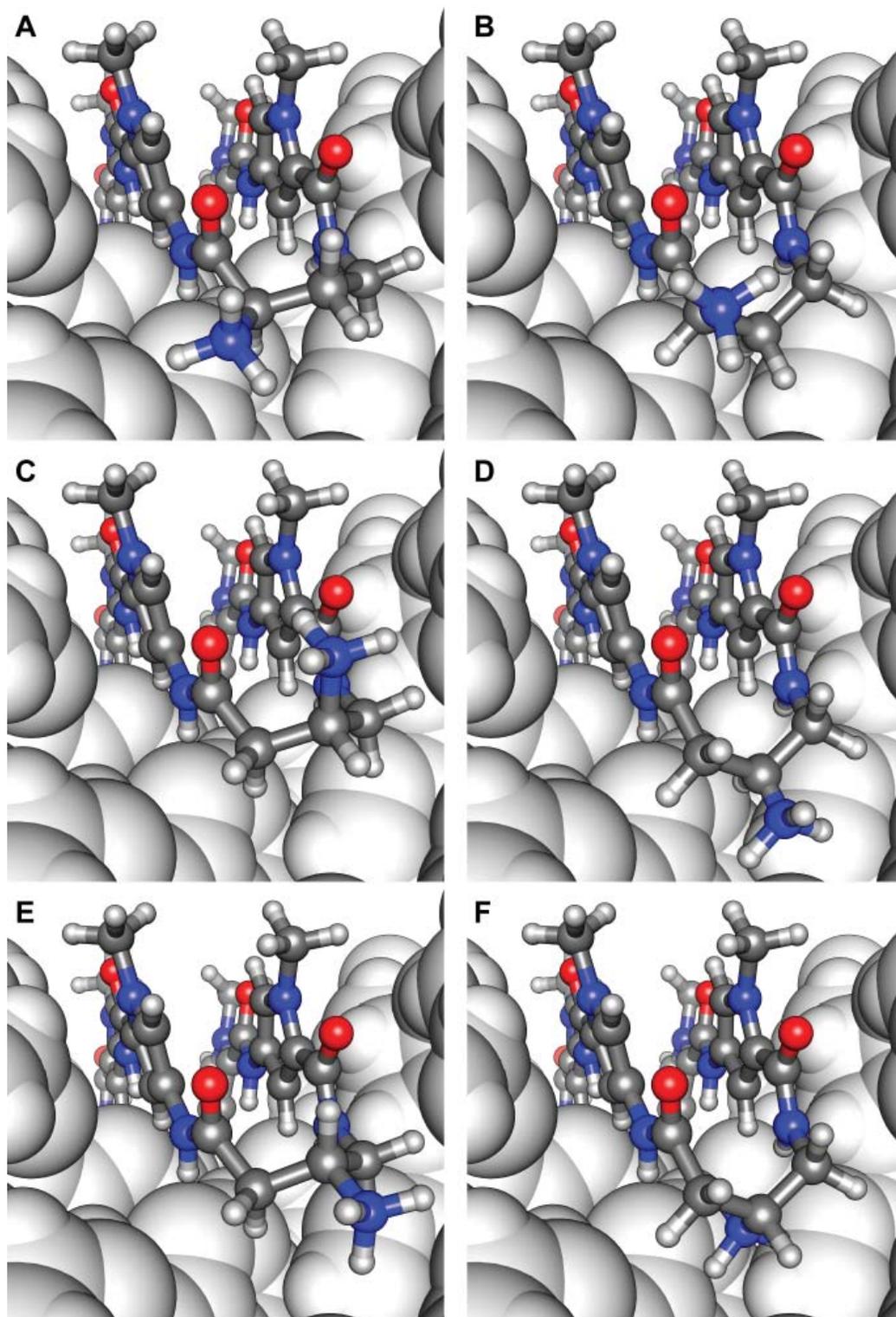
<sup>a</sup> All values reported are derived from at least three melting temperature experiments with standard deviations indicated in parentheses (n.d. = not determined).  $\Delta T_m$  values are given as  $T_m^{(\text{DNA/polyamide})} - T_m^{(\text{DNA})}$ .

have been omitted due to the palindromic core sequence specified by the polyamides. As shown in Table A.1, most  $\gamma$ -turn and (R)- $\alpha$ -amino- $\gamma$ -turn hairpins provided similar  $\Delta T_m$  values for T•A and A•T base pairs. In contrast, significantly lower thermal stabilizations for T•A over A•T base pairs were observed for  $\beta$ -amino- $\gamma$ -turn-linked polyamides targeting DNA sequences 5'-TGGTCA-3'

This is supported by the observation that the presence of fewer continuous Im's improves the DNA affinity of  $\beta$ -amino- $\gamma$ -turns while diminishing the affinity for  $\alpha$ -amino- $\gamma$ -turns, and vice versa. However, illustrative modeling is not sufficient to explain the sequence context dependence of chiral hairpin polyamides, highlighting the pressing need for high-resolution structural studies.

#### A.2.2 Sequence specificity at the turn position

Hairpin polyamides containing the  $\gamma$ -turn have been shown to possess an equal preference for A•T/T•A over G•C/C•G base pairs at the turn position, presumably for steric reasons.<sup>16</sup> Sugiyama's  $\alpha$ -hydroxy- $\gamma$ -turns have been demonstrated to discriminate A•T versus T•A at the turn position.<sup>10c</sup> In order to study the sequence specificity for polyamides **5-16**, we performed melting temperature analyses in the presence of DNA duplexes bearing a T•A base pair at the turn position. Experiments involving hairpins **1-4**

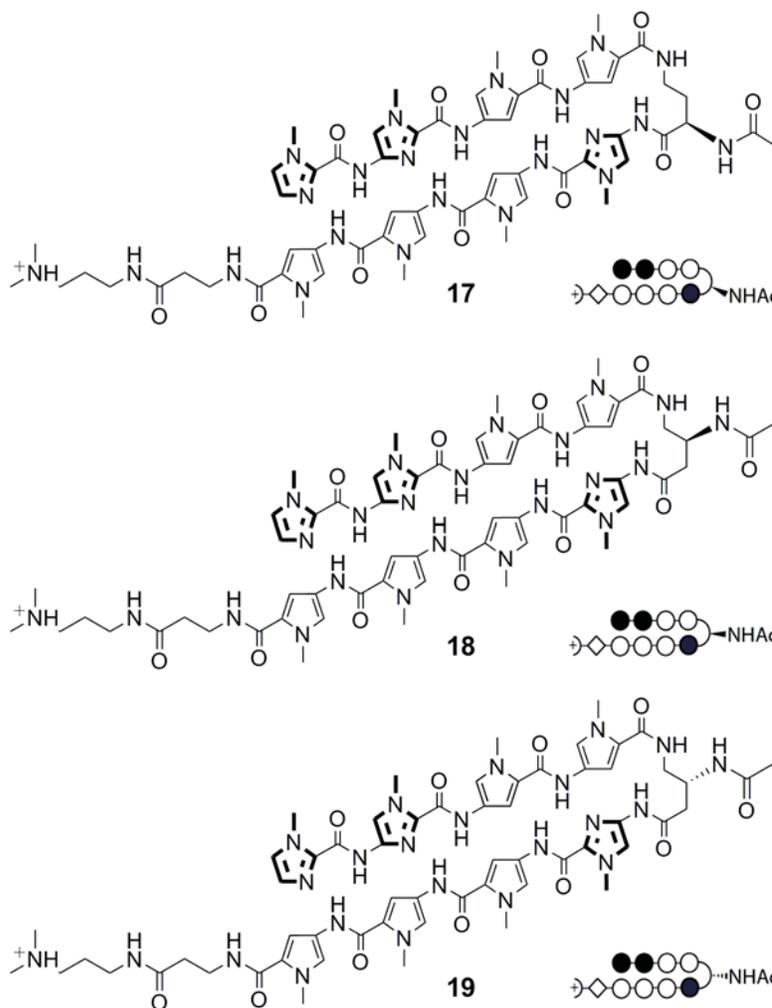


**Figure A.4** Illustrative models of different turn conformations for hairpin polyamides containing the (R)- $\alpha$ -amino- $\gamma$ -turn (A and B), (S)- $\beta$ -amino- $\gamma$ -turn (C and D), and the (R)- $\beta$ -amino- $\gamma$ -turn (E and F) bound to the minor groove of DNA (dark gray = carbons, white = hydrogen, blue = nitrogen, red = oxygen).

(5-8) and 5'-TGGGCA-3' (9-12). Even more diminished duplex stabilizations were observed in presence of C•G and G•C base pairs (see Section A.6 Supplemental Information). These observations suggest that polyamides containing  $\beta$ -amino- $\gamma$ -turns prefer A•T > T•A  $\gg$  C•G > G•C base pairs at the turn position. However, sequence specificity studies by thermal denaturation measurements require binding enthalpies ( $\Delta H_b$ ) of DNA/polyamide complexes in order to determine equilibrium association constants.<sup>14a</sup> One could also imagine using six-ring hairpin polyamides with lower DNA-binding affinities in order to discriminate sequence specificities at the turn position by quantitative DNase I footprint titration methods.

### A.2.3 Acetylated chiral hairpin polyamides

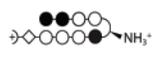
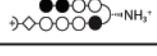
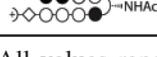
Several approaches have been reported wherein the (R)- $\alpha$ -amino- $\gamma$ -turn was utilized as a position for synthetic modifications of hairpin polyamides.<sup>4,5,17</sup> It has been shown that acetylation of the (R)- $\alpha$ -amine in six-ring hairpin polyamides results in  $\sim$ 15-fold reduced DNA-binding affinity.<sup>10b</sup> To study the tolerance of synthetic modifications for eight-ring polyamides containing the  $\beta$ -amino- $\gamma$ -turns, we examined acetylated hairpins **17-19** by melting temperature analysis (Figure A.5). Indeed, hairpin **17** containing the acetylated (R)- $\alpha$ -amino- $\gamma$ -turn yielded a markedly lower  $\Delta T_m$  value (12.8°C) than nonacetylated



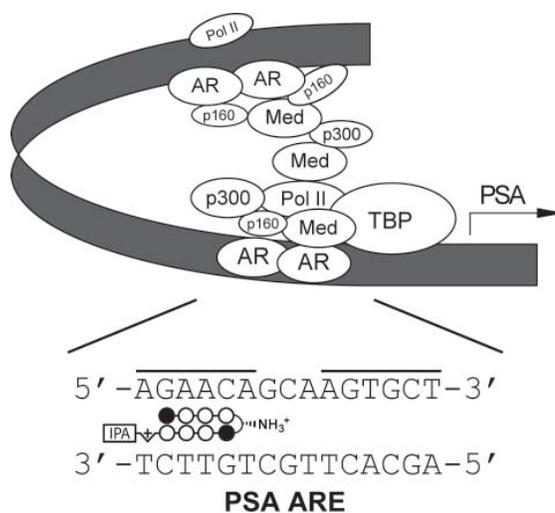
**Figure A.5** Chemical structures and ball-and-stick models of acetylated hairpin polyamides **17-19** targeted to DNA sequence 5'-TGGTCA-3'.

analogue 6 ( $\Delta T_m = 16.9^\circ\text{C}$ , Table A.2). Even more pronounced was the decrease in DNA duplex stabilization for acetylated (S)- $\beta$ -amino- $\gamma$ -turn hairpin **18** leading to a  $\Delta T_m$  value of  $11.7^\circ\text{C}$ . Remarkably, the opposite enantiomer (R)- $\beta$ -**19** resulted in significantly less destabilization ( $\Delta T_m = 17.8^\circ\text{C}$ ). All hairpins lose the positive charge at the turn unit by acetylation. This implies that

**Table A.2** Melting temperatures for DNA complexes containing nonacetylated and acetylated hairpin polyamides targeted to DNA sequence 5'-TGGTCA-3'.<sup>a</sup>

DNA sequence = 5'-CGA TGGTCA AGC-3'		
Polyamides	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$
—	57.2 ( $\pm 0.2$ )	—
 (6)	74.1 ( $\pm 0.3$ )	16.9
 (7)	76.1 ( $\pm 0.2$ )	18.9
 (8)	77.5 ( $\pm 0.3$ )	20.3
 (17)	70.0 ( $\pm 0.1$ )	12.8
 (18)	68.9 ( $\pm 0.2$ )	11.7
 (19)	75.0 ( $\pm 0.1$ )	17.8

<sup>a</sup> All values reported are derived from at least three melting temperature experiments with standard deviations indicated in parentheses.  $\Delta T_m$  values are given as  $T_m^{(\text{DNA/polyamide})} - T_m^{(\text{DNA})}$ .



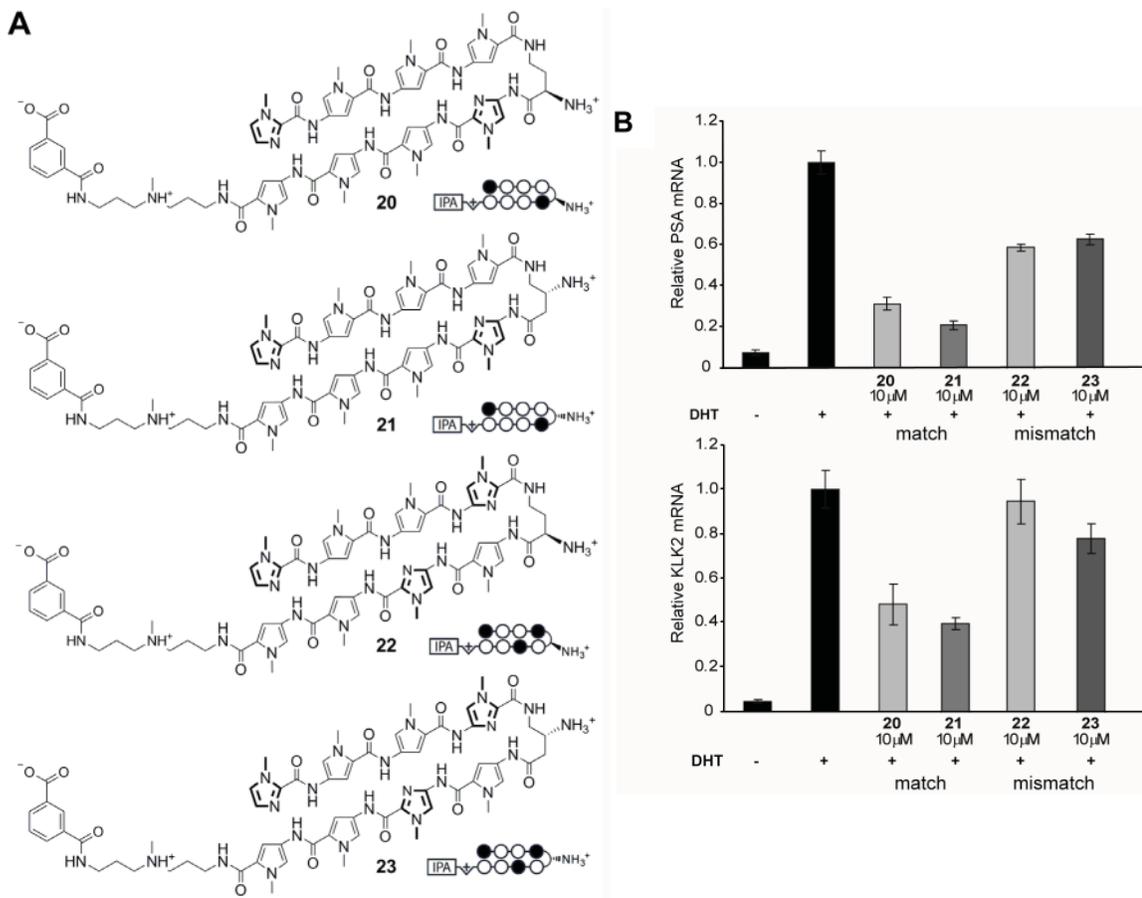
**Figure A.6** Schematic representation of the androgen receptor (AR)-mediated transcription complex with the androgen response element (ARE).

the cationic state of the amine residue is not the only contribution impacting the energetics of the DNA/polyamide complexes, as evidenced by the differences in melting temperatures between hairpins **17-19**. Increased steric demands of the acetylated substituents may also be responsible for differing binding affinities, due to the restricting DNA minor groove and alternate conformations of the  $\gamma$ -turn units (Figure A.4).

#### A.2.4 Biological assay for cell permeability

Hairpin polyamide conjugates bearing the standard (R)- $\alpha$ -amino- $\gamma$ -turn have been shown to modulate the expression of certain gene pathways in living cells by interfering with transcription factor/DNA interfaces.<sup>1</sup> Recently, a hairpin designed to bind DNA sequence 5'-AGAACA-3', found in the androgen response element (ARE), has been demonstrated to inhibit androgen receptor-mediated expression of prostate specific antigen (PSA) in LNCaP cells (Figure A.6).<sup>1b</sup> We utilized this cell culture transcription assay to investigate the cell permeability of (R)- $\beta$ -amino- $\gamma$ -turn hairpins because small structural changes within polyamides can influence nuclear uptake properties.<sup>18</sup> Hairpin polyamide **21** was examined in comparison to the previously used (R)- $\alpha$ -amino- $\gamma$ -turn hairpin

**20** (Figure A.7A). Chiral polyamides **22** and **23**, designed to target different DNA sequences, have been used as controls. Melting temperature analyses for polyamide conjugates **20-23** confirmed the results obtained for hairpins **1-4**, revealing highest DNA-duplex stabilizations for (R)- $\beta$ -amino- $\gamma$ -turn hairpins (see Section A.6 Supplemental Information). The induction of PSA mRNA by dihydrotestosterone (DHT) in LNCaP cells for matched and mismatched polyamides **20-23** was measured by quantitative real-time RT-PCR. As shown in Figure A.7B, hairpin **21** provided



**Figure A.7** (A) Chemical structures and ball-and-stick models of matched and mismatched polyamides **20-23**, targeted to DNA sequence 5'-AGAACA-3'. (B) Inhibition of DHT-induced PSA and KLK2 expression by **20-23** measured by quantitative real-time RT-PCR

significant inhibition of AR-mediated expression of PSA mRNA, KLK2, FKBP5, and TMPRSS2 mRNA (see Section A.6 Supplemental Information) which supports cell-permeable properties for (R)- $\beta$ -amino- $\gamma$ -turn hairpins.

### A.3 Conclusion

Herein we have introduced (R)- and (S)- $\beta$ -amino- $\gamma$ -turn hairpin polyamides. Eight new polyamides targeting different DNA-binding motifs have been synthesized, and their impact on DNA duplex stabilization in relation to hairpins containing the parent  $\gamma$ -turn and the standard (R)- $\alpha$ -amino- $\gamma$ -turn was investigated. It was found that changing the turn unit from the (R)- $\alpha$ -amino- $\gamma$ -turn to either enantiomeric forms of the  $\beta$ -amino- $\gamma$ -turn increases the relative DNA-binding affinity of polyamides targeted to 5'-TGTTCA-3' and 5'-TGGTCA-3' but not to 5'-TGGGCA-3' and 5'-TGGGGA-3' sequences, rendering the impact of  $\beta$ -amino-substituted  $\gamma$ -turns sequence context dependent. Acetylation of the (S)- $\beta$ -amino- $\gamma$ -turn has been demonstrated to significantly impact the DNA-binding affinity but has minimal effect for the (R)- $\beta$ -amino- $\gamma$ -turn, which makes the (R)- $\beta$ -amino residue attractive for synthetic modifications and conjugate design. Upper limits presented by DNase I footprinting titrations of high affinity binders rendered melting temperature analysis a more practical choice for dissecting improvements by structural changes of new turn units in hairpin polyamides. Due to the strong thermal stabilizations, reported for eight-ring hairpin polyamides **1-8** targeted to 5'-TGTTCA-3' and 5'-TGGTCA-3' sequences, it is not unreasonable to estimate that the DNA-binding equilibrium association constants are markedly higher than  $K_a \sim 2 \times 10^{10} \text{ M}^{-1}$ . Biological experiments have demonstrated that (R)- $\beta$ -amino- $\gamma$ -turn hairpins possess biological activity to inhibit AR-mediated gene expression within a human cancer cell line and may have similar uptake properties as polyamides bearing the standard (R)- $\alpha$ -amino- $\gamma$ -turn. Ongoing work is focused on the use of the next generation hairpins in biological investigations as well as turn unit sequence specificity and high-resolution crystallographic studies for DNA/chiral hairpin polyamide complexes. These efforts will be reported in due course.

### A.4 Experimental

#### A.4.1 General

Chemicals and solvents were purchased from Sigma-Aldrich and were used without further purification. Boc- $\gamma$ -Abu-OH was purchased from Novabiochem. (R)-2,4-Fmoc-Dbu(Boc)-OH and Boc- $\beta$ -Ala-PAM resin were purchased from Peptides International. (R)-3,4-Cbz-Dbu(Boc)-OH and (S)-3,4-Cbz-Dbu(Boc)-OH were purchased from Senn Chemicals AG. All DNA oligomers were purchased HPLC purified from Integrated DNA Technologies. Water (18 M $\Omega$ ) was purified using a Millipore MilliQ purification system. The pH of buffers was adjusted using a Beckman 340 pH/temp meter. Analytical HPLC was performed on a Beckman Gold system equipped with a

diode array detector using a Phenomenex Gemini column (5  $\mu\text{m}$  particle size, C18 110A, 250  $\times$  4.6 mm, 5  $\mu\text{m}$ ). Preparative HPLC was performed on a Beckman Gold system equipped with a single-wavelength detector monitoring at 310 nm using a Phenomenex Gemini column (5  $\mu\text{m}$  particle size, C18 110A, 250  $\times$  21.2 mm, 5  $\mu\text{m}$ ). For both analytical and preparative HPLC, solvent A was 0.1% (v/v) aqueous trifluoroacetic acid (TFA) and solvent B was acetonitrile. Solvent gradients were adjusted as needed. Polyamide concentrations were measured in 0.1% (v/v) aqueous TFA on a Hewlett-Packard diode array spectrophotometer "Model 8452 A" and were determined by using an extinction coefficient of 69200  $\text{M}^{-1} \cdot \text{cm}^{-1}$  at  $\lambda_{\text{max}}$  near 310 nm. Matrix-assisted, LASER desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on an Applied Biosystems Voyager DR Pro spectrometer using  $\alpha$ -cyano-4-hydroxycinnamic acid as matrix.

#### A.4.2 Synthesis of polyamides

Polyamide monomers and oligomers were synthesized as described previously.<sup>19</sup> All  $\beta$ -amino- $\gamma$ -turn hairpins were synthesized by performing the following procedure: the polyamide was cleaved from the resin with 3-(dimethylamino)-1-propylamine, purified by preparative HPLC, and characterized by MALDI-TOF MS, UV-vis spectroscopy, and analytical HPLC. A 500 nmol fraction of the Cbz-protected hairpin polyamide was dissolved in a 9:1 mixture (500  $\mu\text{L}$ ) of TFA and trifluoromethanesulfonic acid (TFMSA). After 5 min reaction time, the solution was flash-frozen by liquid  $\text{N}_2$  and overlaid with  $N,N'$ -dimethylformamide (1 mL). The thawed solution was diluted with 20% aqueous acetonitrile (8 mL), purified by preparative HPLC, and characterized by MALDI-TOF MS, UV-vis spectroscopy, and analytical HPLC. Acetylated polyamides **17-19** were synthesized by performing the following procedure: A 500 nmol fraction of the polyamide was dissolved in  $N,N'$ -dimethylformamide (900  $\mu\text{L}$ ) and a 9:1 mixture of pyridine/acetic anhydride (100  $\mu\text{L}$ ) was added. After 5 min reaction time, the solution was diluted with 10% aqueous TFA (8 mL), purified by preparative HPLC, and characterized by MALDI-TOF MS, UV-vis spectroscopy, and analytical HPLC. Polyamide conjugates **20-23** were synthesized as described previously.<sup>20</sup>

Polyamide **1**: MALDI-TOF  $[\text{M} + \text{H}]^+$  calculated for  $\text{C}_{58}\text{H}_{72}\text{N}_{21}\text{O}_{10}^+$  = 1222.6, observed = 1222.7.  
 Polyamide **2**: MALDI-TOF  $[\text{M} + \text{H}]^+$  calculated for  $\text{C}_{58}\text{H}_{73}\text{N}_{22}\text{O}_{10}^+$  = 1237.6, observed = 1237.8.  
 Polyamide **3**: MALDI-TOF  $[\text{M} + \text{H}]^+$  calculated for  $\text{C}_{58}\text{H}_{73}\text{N}_{22}\text{O}_{10}^+$  = 1237.6, observed = 1237.8.  
 Polyamide **4**: MALDI-TOF  $[\text{M} + \text{H}]^+$  calculated for  $\text{C}_{58}\text{H}_{73}\text{N}_{22}\text{O}_{10}^+$  = 1237.6, observed = 1237.8.  
 Polyamide **5**: MALDI-TOF  $[\text{M} + \text{H}]^+$  calculated for  $\text{C}_{57}\text{H}_{71}\text{N}_{22}\text{O}_{10}^+$  = 1223.6, observed = 1223.5.  
 Polyamide **6**: MALDI-TOF  $[\text{M} + \text{H}]^+$  calculated for  $\text{C}_{57}\text{H}_{72}\text{N}_{23}\text{O}_{10}^+$  = 1238.6, observed = 1238.6.

Polyamide **7**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{57}H_{72}N_{23}O_{10}^+ = 1238.6$ , observed = 1238.5.  
 Polyamide **8**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{57}H_{72}N_{23}O_{10}^+ = 1238.6$ , observed = 1238.5.  
 Polyamide **9**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{56}H_{70}N_{23}O_{10}^+ = 1224.6$ , observed = 1224.8.  
 Polyamide **10**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{56}H_{71}N_{24}O_{10}^+ = 1239.6$ , observed = 1239.6.  
 Polyamide **11**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{56}H_{71}N_{24}O_{10}^+ = 1239.6$ , observed = 1239.5.  
 Polyamide **12**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{56}H_{71}N_{24}O_{10}^+ = 1239.6$ , observed = 1239.6.  
 Polyamide **13**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{56}H_{70}N_{23}O_{10}^+ = 1224.6$ , observed = 1224.6.  
 Polyamide **14**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{56}H_{71}N_{24}O_{10}^+ = 1239.6$ , observed = 1239.7.  
 Polyamide **15**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{56}H_{71}N_{24}O_{10}^+ = 1239.6$ , observed = 1239.4.  
 Polyamide **16**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{56}H_{71}N_{24}O_{10}^+ = 1239.6$ , observed = 1239.5.  
 Polyamide **17**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{59}H_{74}N_{23}O_{11}^+ = 1280.6$ , observed = 1280.6.  
 Polyamide **18**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{59}H_{74}N_{23}O_{11}^+ = 1280.6$ , observed = 1280.7.  
 Polyamide **19**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{59}H_{74}N_{23}O_{11}^+ = 1280.6$ , observed = 1280.6.  
 Polyamide **20**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{65}H_{77}N_{22}O_{12}^+ = 1357.7$ , observed = 1357.6.  
 Polyamide **21**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{65}H_{77}N_{22}O_{12}^+ = 1357.6$ , observed = 1357.7.  
 Polyamide **22**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{64}H_{76}N_{23}O_{12}^+ = 1358.6$ , observed = 1358.6.  
 Polyamide **23**: MALDI-TOF  $[M + H]^+$  calculated for  $C_{64}H_{76}N_{23}O_{12}^+ = 1358.6$ , observed = 1358.6.

#### A.4.3 UV Absorption Spectrophotometry

Melting temperature analysis was performed on a Varian Cary 100 spectrophotometer equipped with a thermocontrolled cell holder possessing a cell path length of 1 cm. The buffer for the spectroscopic measurements was chosen to match as closely as possible the conditions of DNase I footprinting experiments. We used 10 mM sodium cacodylate since the temperature dependence of Tris-HCl makes it poorly suited for melting temperature analyses.<sup>14a</sup> A degassed aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0 was used as analysis buffer. DNA duplexes and hairpin polyamides were mixed in 1:1 stoichiometry to a final concentration of 2 μM for each experiment. Prior to analysis, samples were heated to 90 °C and cooled to a starting temperature of 25°C with a heating rate of 5°C/min for each ramp. Denaturation profiles were recorded at λ = 260 nm from 25 to 90°C with a heating rate of 0.5°C/min. The reported melting temperatures were defined as the maximum of the first derivative of the denaturation profile.

#### A.4.4 Molecular Modeling

DNA/polyamide models are based on coordinates derived from NMR structure studies using standard bond length and angles.<sup>3c</sup> The molecular graphics images are nonminimized and have been created by introducing ammonium residues to the appropriate position of the turn unit using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).<sup>21</sup>

#### A.4.5 Measurement of Androgen-Induced PSA mRNA

Experiments were performed as described previously.<sup>1b</sup>

### A.5 Notes and References

1. (a) Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B. Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16768-16773; (b) Nickols, N. G.; Dervan, P. B. Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418-10423; (c) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface. *ACS Chemical Biology* **2007**, *2*, 561-571.
2. (a) Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235; (b) Dervan, P. B.; Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.
3. (a) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science* **1998**, *282*, 111-115; (b) Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C. Structural basis for G•C recognition in the DNA minor groove. *Nat. Struct. Biol.* **1998**, *5*, 104-109; (c) Zhang, Q.; Dwyer, T. J.; Tsui, V.; Case, D. A.; Cho, J. H.; Dervan, P. B.; Wemmer, D. E. NMR structure of a cyclic polyamide-DNA complex. *J. Am. Chem. Soc.* **2004**, *126*, 7958-7966; (d) Puckett J. W.; Muzikar K. A.; Tietjen J.; Warren C. L.; Ansari A. Z.; Dervan P. B. Quantitative Microarray Profiling of DNA-Binding Molecules. *J. Am. Chem. Soc.* **2007**, *129*, 12310-12319.
4. (a) Wurtz, N. R.; Dervan, P. B. Sequence specific alkylation of DNA by hairpin pyrrole-imidazole polyamide conjugates. *Chemistry & Biology* **2000**, *7*, 153-161; (b) Sasaki, S.; Bando, T.; Minoshima, M.; Shimizu, T.; Shinohara, K.; Takaoka, T.; Sugiyama, H. Sequence-specific alkylation of double-strand human telomere repeat sequence by pyrrole-imidazole polyamides with indole linkers. *J. Am. Chem. Soc.* **2006**, *128*, 12162-12168; (c) Tsai, S. M.; Farkas, M. E.; Chou, C. J.; Gottesfeld, J. M.; Dervan, P. B. Unanticipated differences between alpha- and gamma-diaminobutyric acid-linked hairpin polyamide-alkylator conjugates. *Nucleic Acids Res.* **2007**, *35*, 307-316; (d) Minoshima, M.; Bando, T.; Sasaki, S.; Shinohara, K.; Shimizu, T.; Fujimoto, J.; Sugiyama, H. DNA alkylation by pyrrole-imidazole seco-CBI conjugates with an indole linker: sequence-specific DNA alkylation with 10-base-pair recognition through heterodimer formation. *J. Am. Chem. Soc.* **2007**, *129*, 5384-5390.
5. Poulin-Kerstien, A. T.; Dervan, P. B. A two-hit mechanism for pre-mitotic arrest of cancer cell proliferation by a polyamide-alkylator conjugate. *J. Am. Chem. Soc.* **2003**, *125*, 15811-15821.

6. (a) Fechter, E. J.; Dervan, P. B. Allosteric inhibition of protein-DNA complexes by polyamide-intercalator conjugates. *J. Am. Chem. Soc.* **2003**, *125*, 8476-8485; (b) Fechter, E. J.; Olenyuk, B.; Dervan, P. B. Design of a sequence-specific DNA bisintercalator. *Angew. Chem. Int. Ed.* **2004**, *43*, 3591-3594; (c) Fechter, E. J.; Olenyuk, B.; Dervan, P. B. Sequence-specific fluorescence detection of DNA by polyamide-thiazole orange conjugates. *J. Am. Chem. Soc.* **2005**, *127*, 16685-16691.
7. (a) Rucker, V. C.; Foister, S.; Melander, C.; Dervan, P. B. Sequence specific fluorescence detection of double strand DNA. *J. Am. Chem. Soc.* **2003**, *125*, 1195-1202; (b) Chenoweth, D. M.; Viger, A.; Dervan, P. B. Fluorescent sequence-specific dsDNA binding oligomers. *J. Am. Chem. Soc.* **2007**, *129*, 2216-2217.
8. (a) Cohen, J. D.; Sadowski, J. P.; Dervan, P. B. Addressing single molecules on DNA nanostructures. *Angew. Chem. Int. Ed.* **2007**, *46*, 7956-7959; (b) Schmidt, T. L.; Nandi, C. K.; Rasched, G.; Parui, P. P.; Brutschy, B.; Famulok, M.; Heckel, A. Polyamide struts for DNA architectures. *Angew. Chem. Int. Ed.* **2007**, *46*, 4382-4384; (c) Cohen, J. D.; Sadowski, J. P.; Dervan, P. B. Programming multiple protein patterns on a single DNA nanostructure. *J. Am. Chem. Soc.* **2008**, *130*, 402-403.
9. (a) Arndt, H. D.; Hauschild, K. E.; Sullivan, D. P.; Lake, K.; Dervan, P. B.; Ansari, A. Z. Toward artificial developmental regulators. *J. Am. Chem. Soc.* **2003**, *125*, 13322-13323; (b) Kwon, Y.; Arndt, H. D.; Qian, M.; Choi, Y.; Kawazoe, Y.; Dervan, P. B.; Uesugi, M. Small molecule transcription factor mimic. *J. Am. Chem. Soc.* **2004**, *126*, 15940-15941; (c) Hauschild, K. E.; Metzler, R. E.; Arndt, H. D.; Moretti, R.; Raffaele, M.; Dervan, P. B.; Ansari, A. Z. Temperature-sensitive protein-DNA dimerizers. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5008-5013; (d) Stafford, R. L.; Arndt, H. D.; Brezinski, M. L.; Ansari, A. Z.; Dervan, P. B. Minimization of a protein-DNA dimerizer. *J. Am. Chem. Soc.* **2007**, *129*, 2660-2668; (e) Stafford, R. L.; Dervan, P. B. The reach of linear protein-DNA dimerizers. *J. Am. Chem. Soc.* **2007**, *129*, 14026-14033; (f) Xiao, X.; Yu, P.; Lim, H. S.; Sikder, D.; Kodadek, T. A cell-permeable synthetic transcription factor mimic. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865-2868.
10. (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. Hairpin peptide motif. A new class of oligopeptides for sequence-specific recognition in the minor groove of double-helical DNA. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988; (b) Herman, D. M.; Baird, E. E.; Dervan, P. B. Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides in the minor groove. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391; (c) Zhang, W.; Minoshima, M.; Sugiyama, H. Base pair recognition of the stereochemically alpha-substituted gamma-turn of pyrrole/imidazole hairpin polyamides. *J. Am. Chem. Soc.* **2006**, *128*, 14905-14912; (d) Farkas, M. E.; Tsai, S. M.; Dervan, P. B. Alpha-diaminobutyric acid-linked hairpin polyamides. *Bioorg. Med. Chem.* **2007**, *15*, 6927-6936.
11. Hsu, C. F.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C. C.; Dervan, P. B. Completion of a programmable DNA-binding small molecule library. *Tetrahedron* **2007**, 6146-6151.
12. Trauger, J. W.; Dervan, P. B. Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. *Methods Enzymol.* **2001**, *340*, 450-466.
13. For quantitative footprinting experiments, the DNA concentrations of equilibrium mixtures should be at least 10-fold less than the total association constant of the DNA/ligand complex in order to ensure the approximation  $[\text{ligand}]_{\text{free}} = [\text{ligand}]_{\text{total}}$  for numerical analysis. However, the concentration of the labeled DNA fragment specified by the standard DNA/polyamide footprinting protocol is  $\sim 5$  pM.<sup>12</sup> Consequently, DNA-association constants become compressed and hence unreliable for comparison studies at  $K_a$  values  $\geq 2 \times 10^{10} \text{ M}^{-1}$ . (a) Brenowitz, M.; Senear, D. F.;

- Shea, M. A.; Ackers, G. K. Quantitative DNase footprint titration: a method for studying protein-DNA interactions. *Methods Enzymol.* **1986**, *130*, 132-181; (b) Senear, D. F.; Dalma-Weiszhausz, D. D.; Brenowitz, M. Effects of anomalous migration and DNA to protein ratios on resolution of equilibrium constants from gel mobility-shift assays. *Electrophoresis* **1993**, *14*, 704-712.
14. (a) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. Binding of a hairpin polyamide in the minor groove of DNA: Sequence-specific enthalpic discrimination. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8306-8311; (b) Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J. The thermodynamics of polyamide-DNA recognition: hairpin polyamide binding in the minor groove of duplex DNA. *Biochemistry* **1999**, *38*, 2143-2151.
15. (a) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226; (b) Floreancig, P. E.; Swalley, S. E.; Trauger, J. W.; Dervan, P. B. Recognition of the minor groove of DNA by hairpin polyamides containing  $\alpha$ -substituted- $\beta$ -amino acids. *J. Am. Chem. Soc.* **2000**, *122*, 6342-6350.
16. Swalley, S. E.; Baird, E. E.; Dervan, P. B. Effects of gamma-turn and beta-tail amino acids on sequence-specific recognition of DNA by hairpin polyamides. *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.
17. (a) Weyermann, P.; Dervan, P. B. Recognition of ten base pairs of DNA by head-to-head hairpin dimers. *J. Am. Chem. Soc.* **2002**, *124*, 6872-6878; (b) Herman, D. M.; Baird, E. E.; Dervan, P. B. Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Eur. J.* **1999**, *5*, 975-983; (c) Wang, C. C. C.; Dervan, P. B. Sequence-specific trapping of topoisomerase I by DNA binding polyamide-camptothecin conjugates. *J. Am. Chem. Soc.* **2001**, *123*, 8657-8661; (d) Edayathumangalam, R. S.; Weyermann, P.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K. Molecular recognition of the nucleosomal "supergroove". *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6864-6869.
18. (a) Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B. Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12063-12068; (b) Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B. Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res.* **2004**, *32*, 2802-2818.
19. Baird, E. E.; Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
20. Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. Improved nuclear localization of DNA-binding polyamides. *Nucleic Acids Res.* **2007**, *35*, 363-370.
21. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera – A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.

## A.6 Supplemental Information

Construction of Plasmids pCDMF-1 and pCDMF-2: Oligonucleotides were purchased from Integrated DNA Technologies. The plasmids pCDMF-1 and pCDMF-2 were constructed by annealing the oligonucleotides: 5'-AGCTGCGGCTCGAGACGGCTAACCCATCGAGACGGCTAGCCCATCG-AGACGGCTATCCCATCGAGACGGCTACCCCATCGAGAGGATC-3' and 5-GATCGATCCTCT-CGATGGGGTAGCCGTCTCGATGGGATAGCCGTCTCGATGGGGCTAGCCGTCTCGATGGGGTTA-GCCGTCTCGAGCCGC-3'; 5'-AGCTGCGAGACGGCTCGAGACGGCTTGAACATCGAGACGG-CTCGAGACGGCTTGACCATCGAGACGGCTCGAGACGGCTC-3' and 5-GATCGAGCCGTCTCG-AGCCGTCTCGATGGTCAAGCCGTCTCGAGCCGTCTCGATGTTCAAGCCGTCTCGAGCCGTCTCGC-3', respectively, followed by ligation into the BamHI/HindIII restriction fragment of pUC19 using T4 DNA ligase. The plasmid was then transformed into Escherichia coli JM109 competent cells. Ampicillin-resistant white colonies were selected from 25 mL Luria–Bertani (LB) agar plates containing 50 mg/mL ampicillin treated with XGAL and isopropyl-β-D-thiogalactopyranoside (IPTG) solutions and grown overnight at 37°C. Cells were harvested the following day and purification of the plasmid was performed with a Wizard Plus Midiprep DNA purification kit (Promega). DNA sequencing of the plasmid insert was performed by the sequence analysis facility at the California Institute of Technology.

DNase I Footprinting Titrations: Polyamide equilibrations and DNase I footprint titrations were conducted on the 5' end-labeled PCR product of pCDMF-1 and pCDMF-2 according to standard protocols.<sup>12</sup> DNA was incubated with polyamides or water (control) for 12 h at room temperature prior to reaction with DNase I.

**Table A.3** Melting temperatures of DNA/polyamide complexes for all four base pair variations at the turn position of hairpin polyamides.<sup>a</sup>

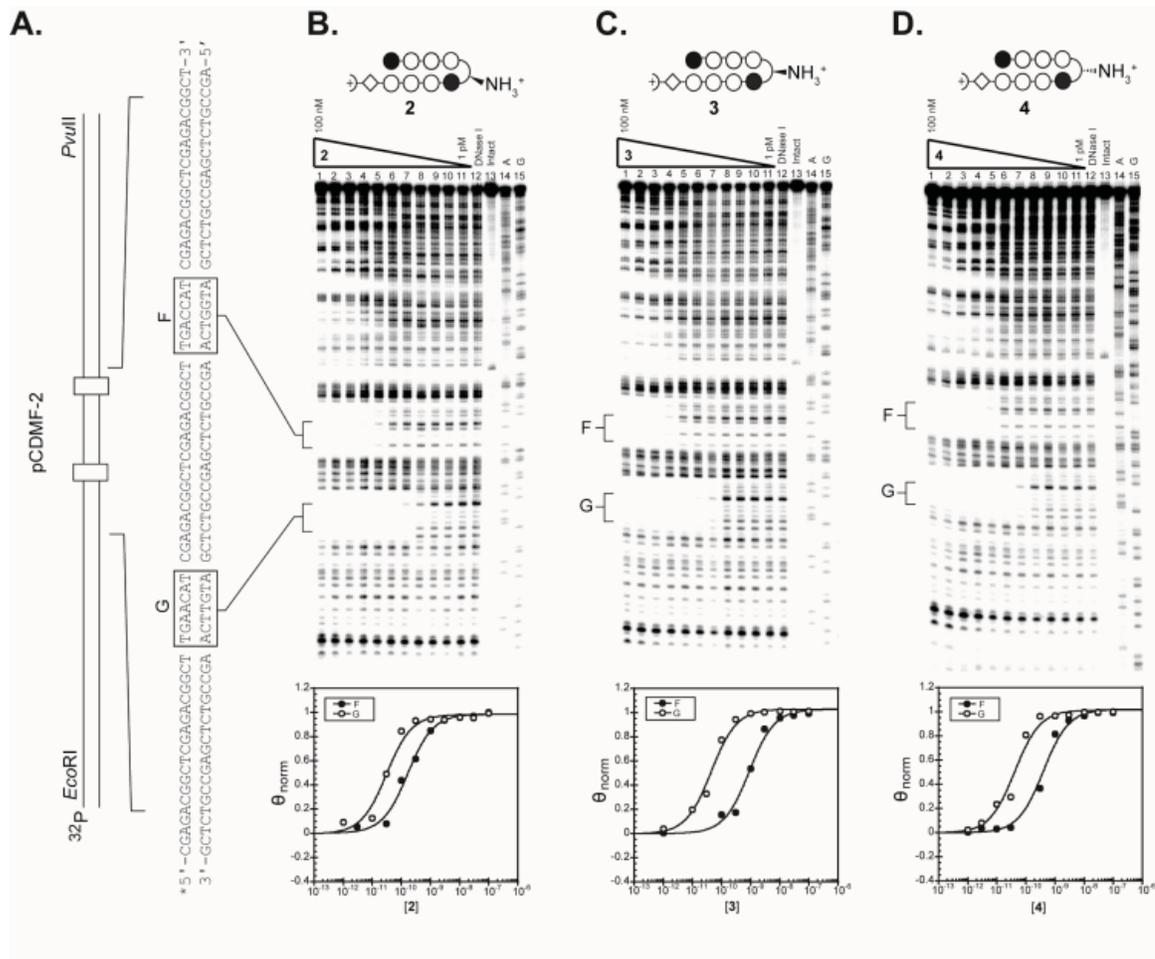
Polyamides	A•T		T•A		C•G		G•C	
	5'-CGA <b>TGGTCA</b> AGC-3'		5'-CGA <b>TGGTCT</b> AGC-3'		5'-CGA <b>TGGTCC</b> AGC-3'		5'-CGA <b>TGGTCG</b> AGC-3'	
	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$
—	57.2 (±0.1)	—	55.8 (±0.1)	—	59.7 (±0.3)	—	60.4 (±0.2)	—
(5)	70.6 (±0.2)	13.4	69.0 (±0.3)	13.2	65.9 (±0.3)	6.2	64.3 (±0.1)	3.9
(6)	74.1 (±0.3)	16.9	72.9 (±0.2)	17.1	67.3 (±0.2)	7.6	64.7 (±0.2)	4.3
(7)	76.1 (±0.2)	18.9	73.2 (±0.1)	17.4	69.7 (±0.3)	10.0	66.1 (±0.2)	5.7
(8)	77.5 (±0.3)	20.3	74.2 (±0.1)	18.4	70.1 (±0.1)	10.4	66.8 (±0.2)	6.4

<sup>a</sup> All values reported are derived from at least three melting temperature experiments with standard deviations indicated in parentheses (n.d. = not determined).  $\Delta T_m$  values are given as  $T_m^{(\text{DNA/polyamide})} - T_m^{(\text{DNA})}$ .

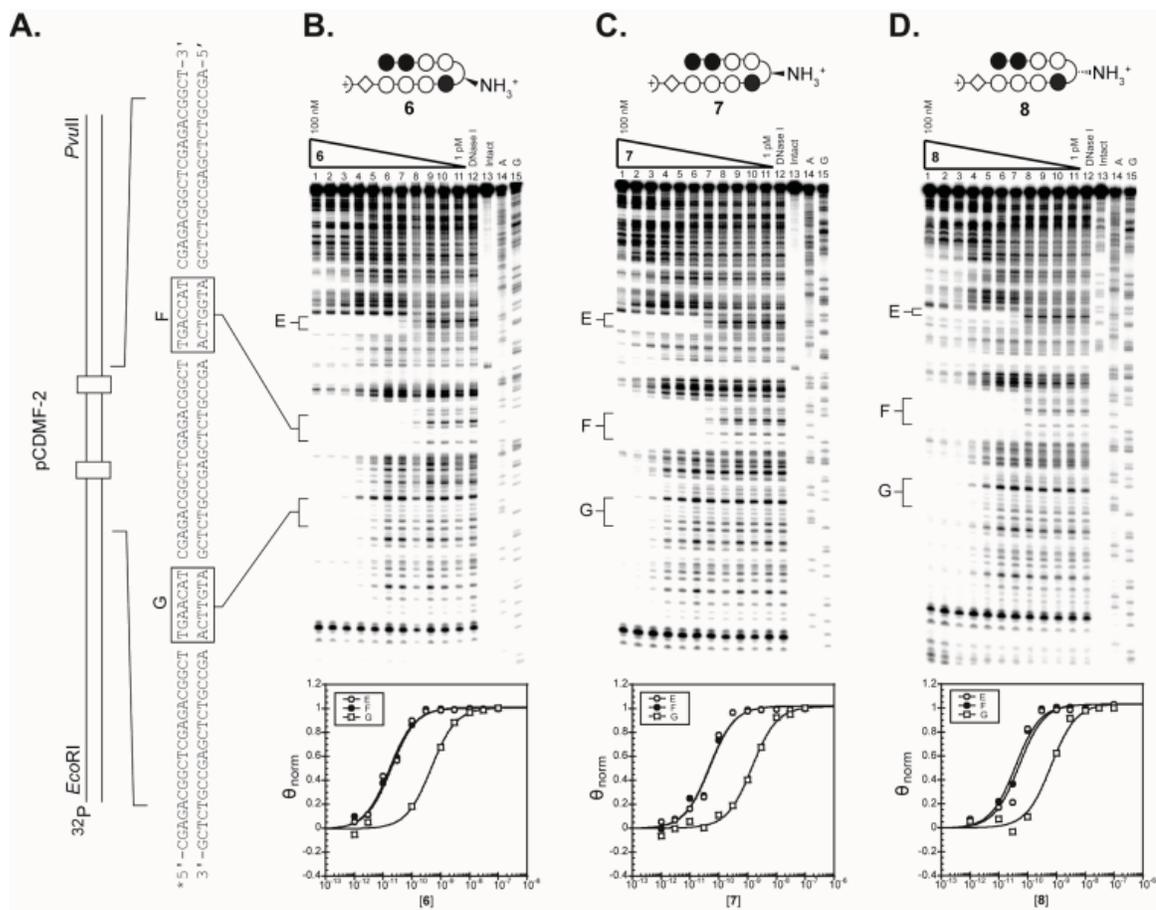
**Table A.4** Equilibrium association constants for hairpin polyamides determined by quantitative DNase I footprint titrations.<sup>a</sup>

Polyamides	A•T		T•A	
	5'-CGA <b>TGTTCA</b> AGC-3'	$K_a / M^{-1}$	5'-CGA <b>TGTTCT</b> AGC-3'	$K_a / M^{-1}$
 (1)		$3.0 (\pm 0.8) \times 10^{10 [b]}$		n. d.
 (2)		$2.6 (\pm 0.6) \times 10^{10}$		n. d.
 (3)		$2.1 (\pm 0.1) \times 10^{10}$		n. d.
 (4)		$2.7 (\pm 0.3) \times 10^{10}$		n. d.
	5'-CGA <b>TGGTCA</b> AGC-3'		5'-CGA <b>TGGTCT</b> AGC-3'	
 (5)		$1.3 (\pm 0.7) \times 10^{10 [b]}$		n. d.
 (6)		$3.1 (\pm 0.5) \times 10^{10}$		n. d.
 (7)		$2.4 (\pm 0.3) \times 10^{10}$		n. d.
 (8)		$2.3 (\pm 0.3) \times 10^{10}$		n. d.
	5'-CGA <b>TGGGCA</b> AGC-3'		5'-CGA <b>TGGGCT</b> AGC-3'	
 (9)		n. d.		n. d.
 (10)		n. d.		$1.5 (\pm 0.2) \times 10^{10}$
 (11)		n. d.		$3.0 (\pm 0.4) \times 10^9$
 (12)		n. d.		$5.9 (\pm 0.9) \times 10^9$
	5'-CGA <b>TGGGGA</b> AGC-3'		5'-CGA <b>TGGGGT</b> AGC-3'	
 (13)		$2.8 (\pm 0.2) \times 10^{7 [b]}$		n. d.
 (14)		n. d.		$6.6 (\pm 1.8) \times 10^9$
 (15)		n. d.		$9.4 (\pm 3.0) \times 10^7$
 (16)		n. d.		$2.1 (\pm 0.6) \times 10^8$

<sup>a</sup> Equilibrium association constants reported are mean values from at least three quantitative DNase I footprint titration experiments. Standard deviations are shown in parentheses. <sup>b</sup> Equilibrium association constants have been reported previously<sup>11</sup> (n. d. = not determined).



**Figure A.8** Quantitative DNase I footprint titration experiments for polyamides **2**, **3**, and **4** on the 285 base pair, 5′ end-labeled PCR product of plasmid pCDMF-2: lanes 1-11, 100 nM, 30 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM, and 1 pM polyamide, respectively; lane 12, DNase I standard; lane 13, intact DNA; lane 14, A reaction; lane 15, G reaction.



**Figure A.9** Quantitative DNase I footprint titration experiments for polyamides **6**, **7**, and **8** on the 285 base pair, 5' end-labeled PCR product of plasmid pCDMF-2: lanes 1-11, 100 nM, 30 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM, and 1 pM polyamide, respectively; lane 12, DNase I standard; lane 13, intact DNA; lane 14, A reaction; lane 15, G reaction.

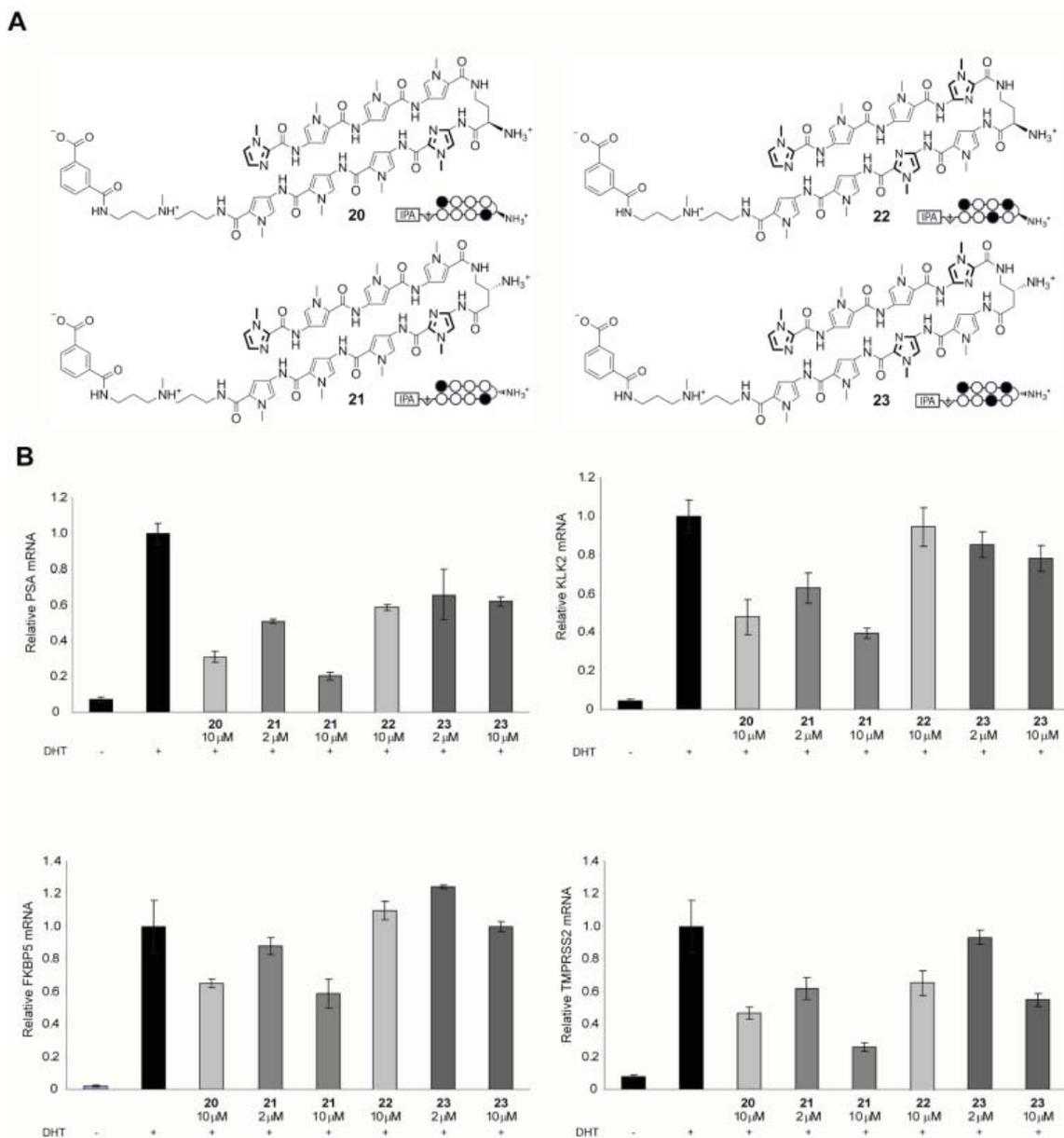




**Table A.5** Melting temperatures of polyamides targeted to DNA-sequence 5'-AGAACA-3' in complex with DNA.<sup>a</sup>

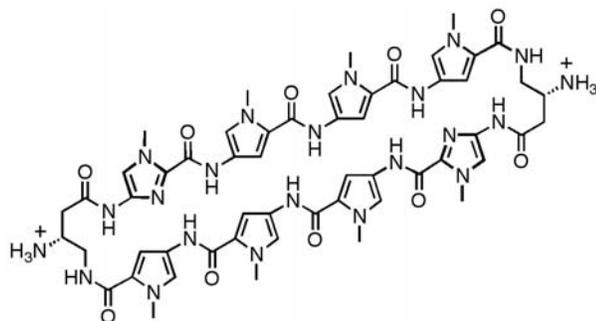
DNA sequence = 5'-TTGC <b>AGAACA</b> GCAA-3'		
Polyamides	$T_m / ^\circ\text{C}$	$\Delta T_m / ^\circ\text{C}$
—	60.1 ( $\pm 0.2$ )	—
 <b>(20)</b>	74.4 ( $\pm 0.2$ )	14.3
 <b>(21)</b>	76.3 ( $\pm 0.2$ )	16.2
 <b>(22)</b>	64.6 ( $\pm 0.1$ )	4.5
 <b>(23)</b>	66.9 ( $\pm 0.1$ )	6.8

<sup>a</sup> All values reported are derived from at least three melting temperature experiments with standard deviations indicated in parentheses.  $\Delta T_m$  values are given as  $T_m^{(\text{DNA/polyamide})} - T_m^{(\text{DNA})}$ .



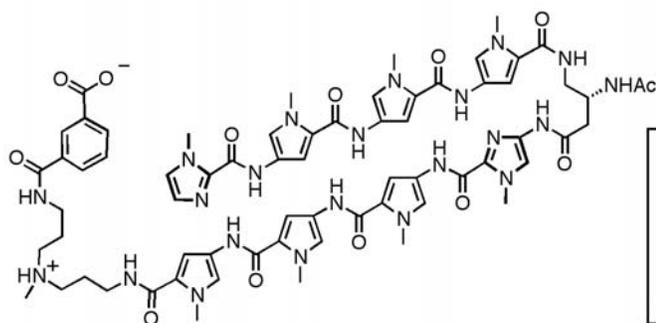
**Figure A.12** A) Chemical structures and ball-and-stick models of matched and mismatched polyamides **20-23**, respectively, targeted to 5'-AGAACA-3'. B) Inhibition of DHT-induced PSA, KLK2, FKBP5, and TMPRSS2 expression by **20-23** measured by quantitative real-time RT-PCR.

**Appendix B: Aprelica ADMET Report**  
(Supplemental Information Continued from Chapter 3)



*Cyclic polyamide 1*

Denoted as **DMC2-239** in ADMET data tables (Table 3.2-3.7) and in full ADMET report located in Appendix B



*Hairpin polyamide 5*

Denoted as **DH-V-88** in ADMET data tables (Table 3.2-3.7) and in full ADMET report located in Appendix B

**Figure B.1** Polyamides **1** and **5** were subjected to preclinical ADMET testing by contract service at Apredica (Watertown, MA). Shown in Chapter 3 of this thesis (Table 3.2-3.7) are summaries of the ADMET results taken directly from the final report provided by Apredica. The full ADMET report, which includes experimental conditions, is contained on the following pages.



**Apredica Study Number: CIT-001**

**ADMET Properties of Test Agents**

**Final Report**



**Sponsor:** California Institute of Technology  
Division of Chemistry and Chemical Engineering  
1200 E California Blvd; MC 164-30  
Pasadena, CA 91125 (USA)

**Sponsor's Representative** Daniel Harki, Ph.D.  
Phone: (626) 395-6032  
Email: harki@caltech.edu

**Test Facilities:** Apredica  
313 Pleasant Street  
Watertown, MA 02472  
Telephone Number: 617.812.1911

**Study Director** Jon Gilbert  
In Vitro ADMET, Bioanalysis  
617.812.1911x122  
jon@apredica.com

Approved: \_\_\_\_\_

Date: \_\_\_\_\_

12/2/08

CONFIDENTIAL

## TABLE OF CONTENTS

1	Objective .....	3
1.1	Regulatory Guidelines .....	3
2	Test Articles .....	3
3	Test Methods .....	4
3.1	Analytical Methods .....	4
3.1.1	<i>Method development</i> .....	4
3.1.2	<i>Analysis</i> .....	4
3.2	In vitro ADME-Tox Experimental Conditions .....	4
3.2.1	<i>Caco-2 monolayer permeability experimental conditions</i> .....	4
3.2.2	<i>Cytotoxicity experimental conditions</i> .....	4
3.2.3	<i>Fluorescent cytochrome P450 inhibition experimental conditions</i> .....	5
3.2.4	<i>Microsomal stability experimental conditions</i> .....	5
3.2.5	<i>Plasma stability experimental conditions</i> .....	5
3.2.6	<i>Plasma protein binding experimental conditions</i> .....	5
3.2.7	<i>hERG FastPatch experimental conditions</i> .....	5
4	Results.....	6
4.1	Analytical .....	6
4.1.1	<i>Method development</i> .....	6
4.2	In vitro ADME-Tox Summary .....	7
4.2.1	<i>Caco-2 permeability summary</i> .....	7
4.2.2	<i>Cytotoxicity summary</i> .....	7
4.2.3	<i>Fluorescent Cyp IC50 summary</i> .....	7
4.2.4	<i>hERG FastPatch summary</i> .....	7
4.2.5	<i>Microsomal intrinsic clearance summary</i> .....	8
4.2.6	<i>Plasma half-life summary</i> .....	8
4.2.7	<i>Plasma protein binding summary</i> .....	8
4.3	In vitro ADME-Tox Individual Data .....	10
4.3.1	<i>Caco-2 permeability individual data</i> .....	10
4.3.2	<i>Cytotoxicity individual data</i> .....	10
4.3.3	<i>hERG FastPatch individual data</i> .....	11
4.3.4	<i>Fluorescent cytochrome P450 inhibition individual data</i> .....	13
4.3.5	<i>Microsomal intrinsic clearance individual data</i> .....	16
4.3.6	<i>Plasma half-life individual data</i> .....	19
4.3.7	<i>Plasma protein binding individual data</i> .....	22
5	References .....	23
6	Storage and Retention of Records .....	24
7	Appendices .....	25
7.1	Appendix A. Standard Apremica Methods .....	25
7.2	Appendix B. Sample Spectra and Chromatograms of the Test Agents .....	29

## 1 Objective

The objective of this study is to determine the ADMET properties of test agents.

### 1.1 Regulatory Guidelines

This study was not conducted under US FDA Good Laboratory Practice Regulations (GLPs). Standard operating procedures of Aprelica were used throughout the study.

## 2 Test Articles

<b>Aprelica ID</b>	<b>Client ID</b>	<b>Physical Form</b>	<b>Submitted FW</b>	<b>Parent MW</b>	<b>Stock solutions</b>
CIT-001-01	DH-V-88	Solid	1399	1399	50 mM DMSO
CIT-001-02	DMC2-239	Solid	1407.26	1178.5	50 mM DMSO

Test agent powders were stored at -20 °C. Stock solutions were stored at -20 °C.

### 3 Test Methods

Testing was performed at Apredica in Watertown, MA.

#### 3.1 Analytical Methods

##### 3.1.1 Method development

The signal was optimized for each compound by ESI positive or negative ionization mode. A MS2 scan was used to identify the precursor ion and a product ion analysis was used to identify the best fragment for analysis and to optimize the collision energy. An ionization ranking was assigned indicating the compound's ease of ionization.

##### 3.1.2 Analysis

Samples were analyzed by LC/MS/MS using either an Agilent 6410 mass spectrometer coupled with an Agilent 1200 HPLC and a CTC PAL chilled autosampler, all controlled by MassHunter software (Agilent), or an ABI2000 mass spectrometer coupled with an Agilent 1100 HPLC and a CTC PAL chilled autosampler, all controlled by Analyst software (ABI). After separation on a C18 reverse phase HPLC column (Agilent, Waters, or equivalent) using an acetonitrile-water gradient system, peaks were analyzed by mass spectrometry (MS) using ESI ionization in MRM mode.

#### 3.2 In vitro ADME-Tox Experimental Conditions

Additional protocol details are given in Appendix A.

##### 3.2.1 Caco-2 monolayer permeability experimental conditions

Apredica ID	Client ID	Test conc.	Assay Time	Direction	Reference compounds	Analytical method
CIT-001-01	DH-V-88			A->B	warfarin	
CIT-001-02	DMC2-239	10 $\mu$ M	2 hr	B->A	ranitidine	LC/MS/MS

##### 3.2.2 Cytotoxicity experimental conditions

Apredica ID	Client ID	Test conc.	Assay time	Cell lines	Readout	Reference compound	Analytical method
		100, 40, 16, 6.4, 2.6, 1.0, 0.4, 0.16, 0.07					
CIT-001-01	DH-V-88						fluorescent plate reader
CIT-001-02	DMC2-239	$\mu$ M	48 hr	HepG2 NIH/3T3	Neutral red	chlorpromazine propranolol	

CONFIDENTIAL

### 3.2.3 Fluorescent cytochrome P450 inhibition experimental conditions

Apredica ID	Client ID	Test conc.	Cyp assays	Reference compound	Analytical method
		50, 16.7, 5.6, 1.9, 0.6, 0.2, 0.07, 0.02 $\mu$ M	Cyp1A2/CEC Cyp2C8/DBP Cyp2C9/DBF Cyp2C19/DBF Cyp2D6/AMMC Cyp3A4/DBF Cyp3A4/BFC	$\alpha$ -naphthoflavone ketoconazole sulphaphenazole tranylcypromine quinidine ketoconazole ketoconazole	fluorescent plate reader
CIT-001-01	DH-V-88				
CIT-001-02	DMC2-239				

### 3.2.4 Microsomal stability experimental conditions

Apredica ID	Client ID	Test conc.	Micro-some source	Protein conc.	Incub-ation	Ref. comp.	Analytical method
					0, 10, 20, 40, and 60 min		
CIT-001-01	DH-V-88		Human	0.3 mg/mL	37 °C	verapamil	
CIT-001-02	DMC2-239	5 $\mu$ M	and rat			warfarin	LC/MS/MS

### 3.2.5 Plasma stability experimental conditions

Apredica ID	Client ID	Test conc.	Plasma source	Incub-ation	Reference compounds	Analytical method
				0, 15, 30, 60, 120 min		
CIT-001-01	DH-V-88		Human	37 °C		
CIT-001-02	DMC2-239	10 $\mu$ M	and rat		propranolol	LC/MS/MS

### 3.2.6 Plasma protein binding experimental conditions

Apredica ID	Client ID	Test conc	Plasma species	Incub-ation	Sep. method	Ref. compound	Analytical method
CIT-001-01	DH-V-88						
CIT-001-02	DMC2-239	10 $\mu$ M	Human and rat	4 hr 37 °C	equilibrium dialysis	warfarin, atenolol	LC/MS/MS

### 3.2.7 hERG FastPatch experimental conditions

Apredica ID	Client ID	Test conc	Medium	Incub-ation	Ref. comp.	Analytical method
		100, 30, 10, 3, 1, 0.3, 0.1, 0.03 $\mu$ M	HEPES-aspartate buffer	5 min ambient temp.		
CIT-001-01	DH-V-88					
CIT-001-02	DMC2-239				E-4031	electro-physiology

## 4 Results

### 4.1 Analytical

#### 4.1.1 Method development

Client ID	MW	Polarization	Precursor m/z	Product m/z	Collision energy (V)	Ionization classification <sup>a</sup>
DH-V-88	1399	pos	700.2	231	26	2
DMC2-239	1178.5	pos	590.1	372	20	2

<sup>a</sup>Ionization classification:  
1 = Highly ionizable  
2 = Intermediately ionizable  
3 = Poorly ionizable

The full scan mass spectrum, the product ion spectrum, and a sample chromatogram are shown in Appendix B.

CONFIDENTIAL

## 4.2 In vitro ADME-Tox Summary

### 4.2.1 Caco-2 permeability summary

Client ID	test conc (µM)	Assay duration (hr)	mean A->B $P_{app}^a$ ( $10^{-6}$ cm s <sup>-1</sup> )	mean A->B $P_{app}^a$ ( $10^{-6}$ cm s <sup>-1</sup> )	Asymmetry ratio <sup>b</sup>	comment
Warfarin	50	2	35.4	7.9	0.2	high permeability control
Ranitidine	50	2	1.4	2.4	1.7	low permeability control
DH-V-88	10	2	ND	0.11	UD	
DMC2-239	10	2	ND	ND	ND	

<sup>a</sup>Apparent permeability<sup>b</sup> $P_{app}(B \rightarrow A) / P_{app}(A \rightarrow B)$ 

ND = no compound detected in receiver solution

### 4.2.2 Cytotoxicity summary

Client ID	Cell line	IC <sub>50</sub> (µM)	comment
Chlorpromazine	HepG2	13	Higly cytotoxic control
Propranolol	HepG2	80	Low cytotoxic control
DH-V-88	HepG2	>100	
DMC2-239	HepG2	>100	

### 4.2.3 Fluorescent Cyp IC<sub>50</sub> summary

Client ID	IC <sub>50</sub> (µM)						
	Cyp1A2 / CEC	Cyp2C8/D BF	Cyp2C9 / DBF	Cyp2C19 / DBF	Cyp2D6 / AMMC	Cyp3A4 / BFC	Cyp3A4 / DBF
Controls	0.2 α-naphthoflavone	2.3 ketoconazole	1.1 sulpha-phenazole	5.6 tranyl-cyromine	0.05 quinindine	1.26 ketoconazole	1.26 ketoconazole
DH-V-88	>50	>50	>50	>50	>50	47.6	>50
DMC2-239	>50	>50	>50	>50	>50	37.7	>50

### 4.2.4 hERG FastPatch summary

Client ID	IC <sub>50</sub> (µM)	comment
E-4031	99% at 0.5 µM	positive control
DH-V-88	>100	*
DMC2-239	>100	*

\*The solubility limit for this experiment, as determined by vehicle controls, was 17.3 x 10<sup>3</sup> LSU (horizontal black line). Based on the data obtained, there may be solubility issues for both test articles at 30 and 100 µM in our physiological saline solution (HB-PS, 0.3%DMSO). Precipitation of DH-V-88 at 100 µM was visible to the naked eye.

CONFIDENTIAL

## 4.2.5 Microsomal intrinsic clearance summary

Client ID	test conc (µM)	test species	NADPH-dependent CL <sub>int</sub> <sup>a</sup> (µl min <sup>-1</sup> mg <sup>-1</sup> )	NADPH-dependent T <sub>1/2</sub> <sup>b</sup> (min)	NADPH-free CL <sub>int</sub> <sup>a</sup> (µl min <sup>-1</sup> mg <sup>-1</sup> )	NADPH-free T <sub>1/2</sub> <sup>b</sup> (min)	comment
Verapamil	1	Human	411.3	5.6	0.6	>180	metabolized control
Verapamil	1	Rat	2276	1	0.0	>180	metabolized control
Warfarin	1	Human	0.0	>180	0.0	>180	non-metabolized control
Warfarin	1	Rat	0.0	>180	0.0	>180	non-metabolized control
DH-V-88	5	Human	0.0	>180	0.0	>180	
DH-V-88	5	Rat	0.0	>180	0.0	>180	
DMC2-239	5	Human	0.0	>180	0.0	>180	
DMC2-239	5	Rat	0.0	>180	0.0	>180	

<sup>a</sup>Microsomal Intrinsic Clearance<sup>b</sup>Half-life

## 4.2.6 Plasma half-life summary

Compound	test conc (µM)	medium	T <sub>1/2</sub> (min)	Fraction remaining, max time (%)	comment
Propranolol	10.0	Human Plasma	35.5	5.8%	control
		Rat Plasma	149.0	51.6%	
DH-V-88	10.0	Human Plasma	>120	95.6%	
		Rat Plasma	>120	94.0%	
DMC2-239	10.0	Human Plasma	>120	124.5%	
		Rat Plasma	>120	120.3%	

<sup>a</sup>Half-life

## 4.2.7 Plasma protein binding summary

Client ID	test conc (µM)	Assay duration	Species	Mean free fraction (%)	comment
Warfarin	10	4 hr	Human	0.73%	high binding control
Warfarin	10	4 hr	Rat	5.47%	high binding control
Atenolol	10	4 hr	Human	76.2%	low binding control
Atenolol	10	4 hr	Rat	84.7%	low binding control
DH-V-88	10	4 hr	Human	0.0015%	
DH-V-88	10	4 hr	Rat	0.0016%	
DMC2-239	10	4 hr	Human	0.0000%	
DMC2-239	10	4 hr	Rat	0.0040%	



CONFIDENTIAL

### 4.3 In vitro ADME-Tox Individual Data

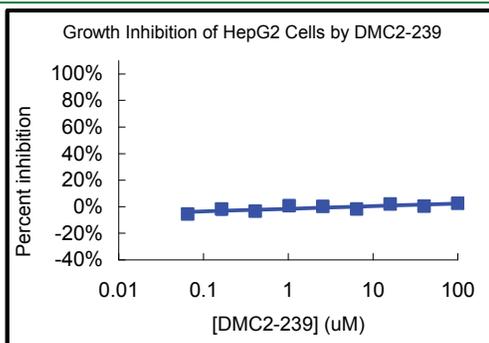
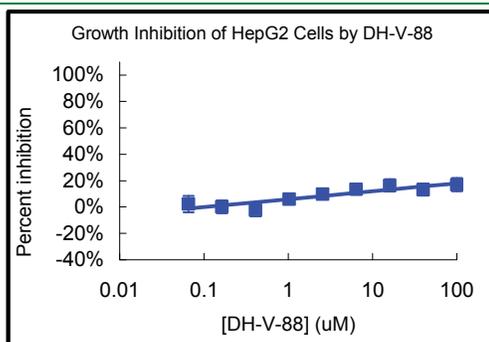
#### 4.3.1 Caco-2 permeability individual data

<i>Client ID</i>	<i>test conc (μM)</i>	<i>direction</i>	<i>value</i>	<i>1<sup>st</sup></i>	<i>2<sup>nd</sup></i>	<i>mean</i>	<i>comment</i>
DH-V-88	10	A->B	dQ/dt <sup>a</sup>	ND	ND	ND	
		A->B	C <sub>0</sub> <sup>b</sup>	1.1	1.0	1.1	
		B->A	dQ/dt <sup>a</sup>	8.6E-08	1.9E-08	5.3E-08	
		B->A	C <sub>0</sub> <sup>b</sup>	1.4	1.4	1.4	
DMC2-239	10	A->B	dQ/dt <sup>a</sup>	ND	ND	ND	
		A->B	C <sub>0</sub> <sup>b</sup>	0.05	0.02	3.5E-02	
		B->A	dQ/dt <sup>a</sup>	ND	ND	ND	
		B->A	C <sub>0</sub> <sup>b</sup>	0.1	0.1	8.4E-02	

<sup>a</sup>rate of test agent permeation, area units/sec

<sup>b</sup>initial concentration (area units/cm<sup>3</sup>)

#### 4.3.2 Cytotoxicity individual data



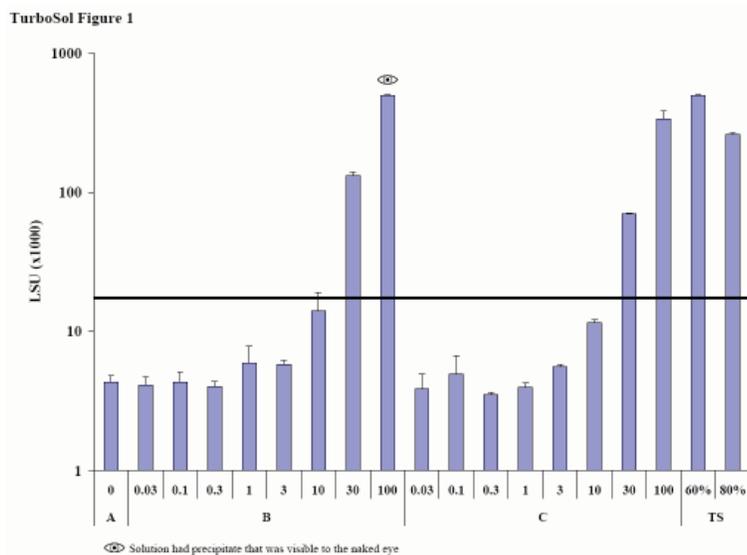
CONFIDENTIAL

## 4.3.3 hERG FastPatch individual data

Test Article ID	IC50 (µM)	Conc. (µM)	Mean % hERG Inhibition	Standard Deviation	Standard Error	n	Individual Data Points (% Inhibition)
DH-188	> 100	0.03	4.6	3.8	2.7	2	7.2
							1.9
		0.1	5.6	3.1	2.2	2	7.8
							3.4
		0.3	1.9	4.0	2.9	2	-1.0
							4.7
		1	9.3	3.1	2.2	2	7.1
							11.5
		3	2.0	4.8	3.4	2	-1.4
							5.4
		10	3.1	2.3	1.6	2	1.6
							4.7
		30	1.2	0.7	0.5	2	1.6
							0.7
100	0.9	1.4	1.0	2	-0.1		
					1.9		

Test Article ID	IC50 (µM)	Conc. (µM)	Mean % hERG Inhibition	Standard Deviation	Standard Error	n	Individual Data Points (% Inhibition)
DMG2	> 100	0.03	-1.2	0.1	0.1	2	-1.1
							-1.3
		0.1	2.1	0.4	0.3	2	2.3
							1.8
		0.3	-4.1	1.2	0.9	2	-3.2
							-4.9
		1	-3.7	3.3	2.4	2	-1.3
							-6.1
		3	-0.7	1.6	1.1	2	-1.9
							0.4
		10	-2.2	2.8	2.0	2	-4.2
							-0.2
		30	5.5	1.4	1.0	2	4.5
							6.5
100	9.0	2.1	1.5	2	7.5		
					10.5		

CONFIDENTIAL

**TurboSol Table 1**

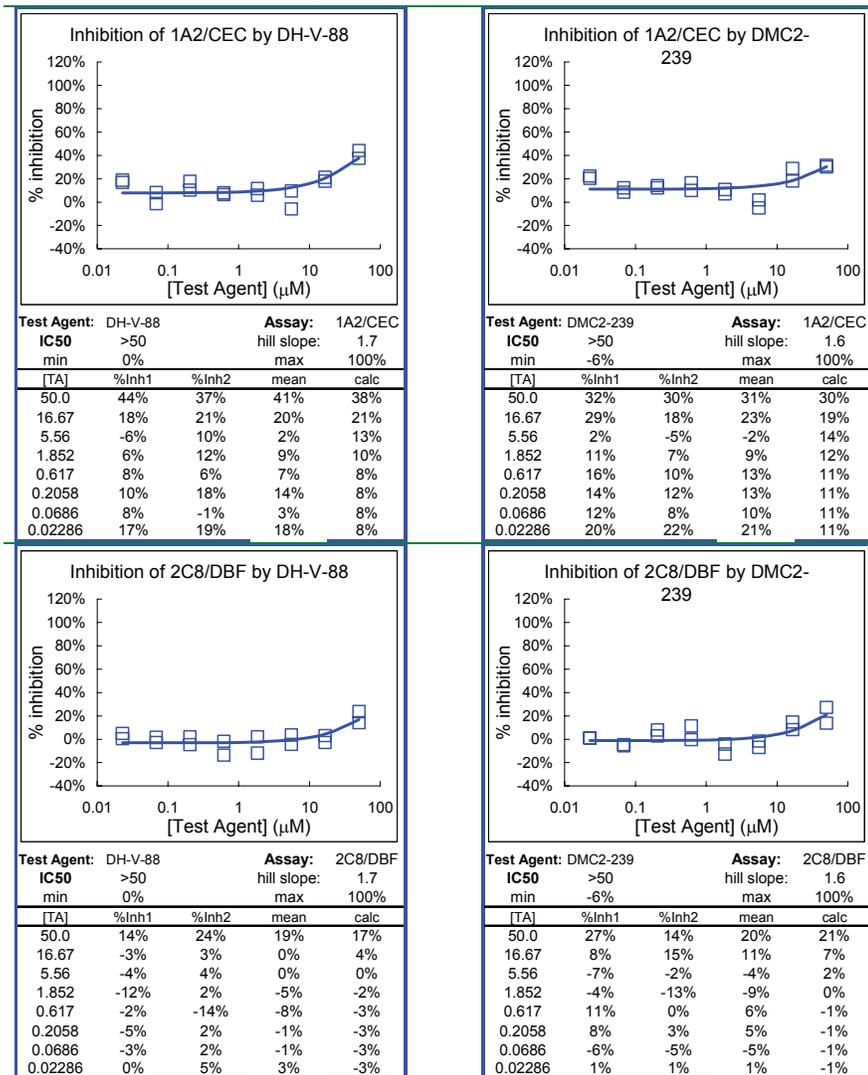
Compound	Figure Legend	Conc. (μM)	n	Average LSU (x1000)	Standard Deviation
Vehicle	A	0	11	6.1	6.2
88	B	0.03	3	4.1	0.6
		0.1	3	4.3	0.8
		0.3	3	4.0	0.4
		1	3	6.0	1.9
		3	3	5.8	0.5
		10	3	14.1	4.9
		30	3	131.9	7.4
		100	3	495.9	7.1
239	C	0.03	3	3.9	1.1
		0.1	3	5.0	1.7
		0.3	3	3.5	0.1
		1	3	3.9	0.3
		3	3	5.6	0.2
		10	3	11.6	0.5
		30	3	70.3	0.2
		100	3	335.9	46.3
Transmittance Standard	TS	60%	3	495.3	7.3
		80%	3	259.0	9.3

TS: Transmittance standard. % indicates percent light transmitted.

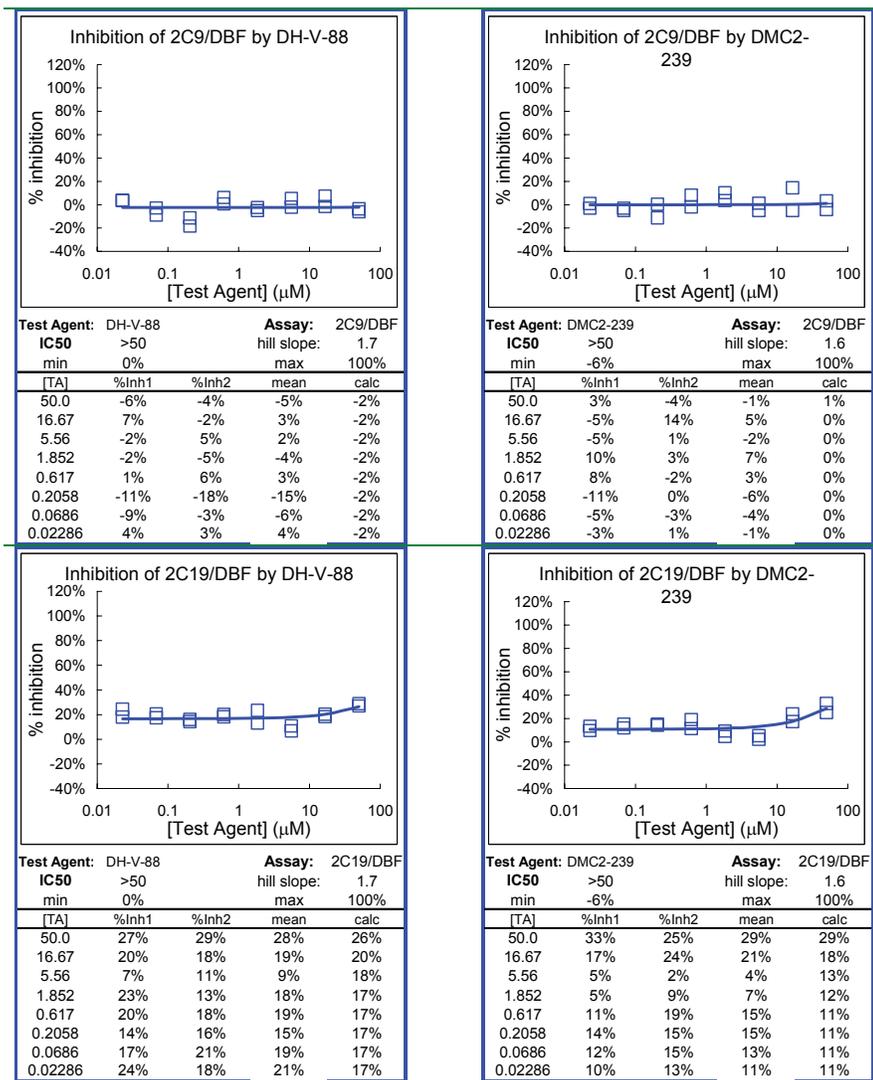
LSU: Light scatter unit

CONFIDENTIAL

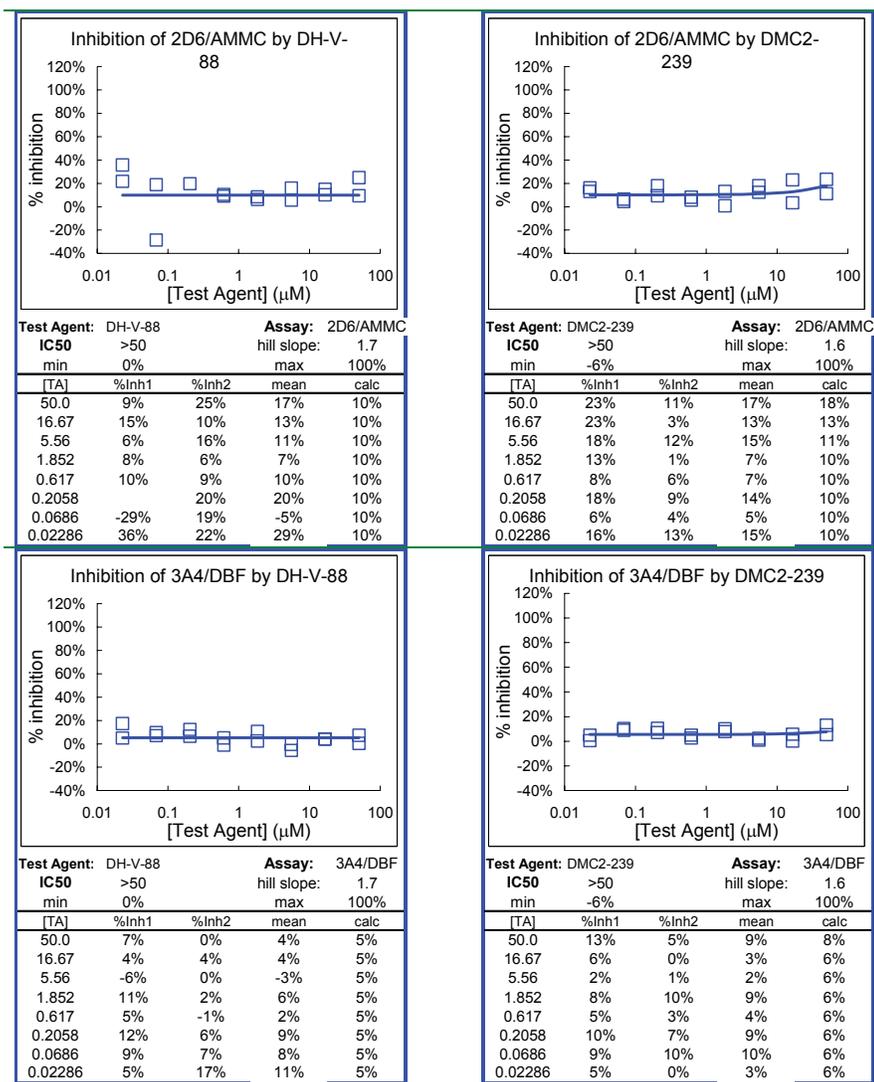
## 4.3.4 Fluorescent cytochrome P450 inhibition individual data



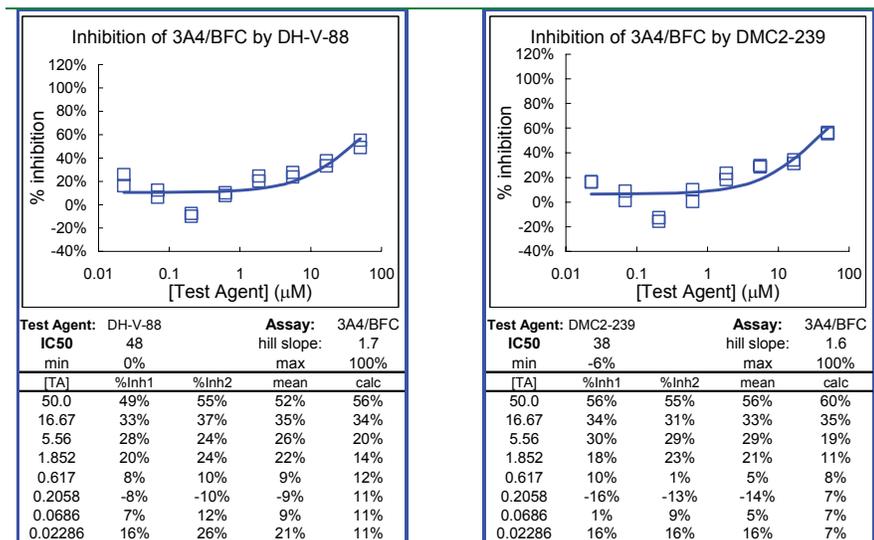
CONFIDENTIAL



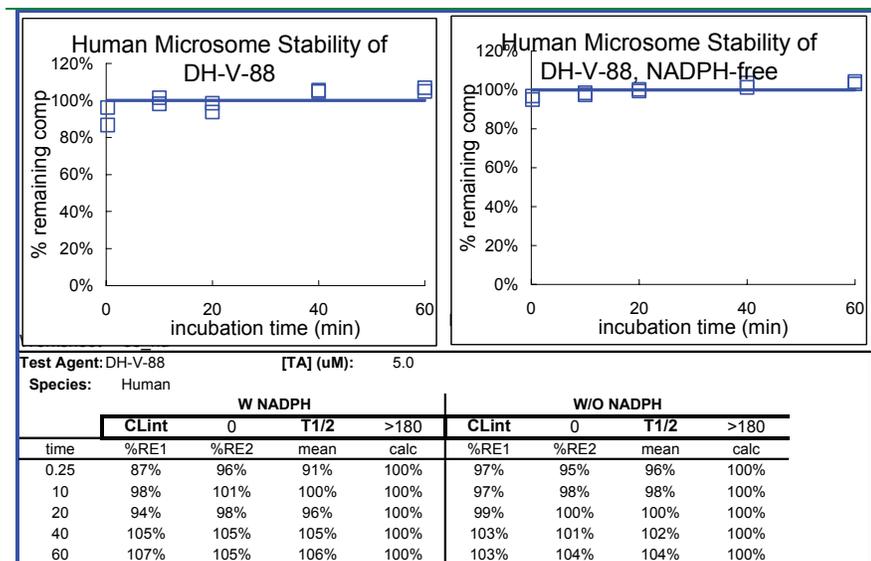
CONFIDENTIAL



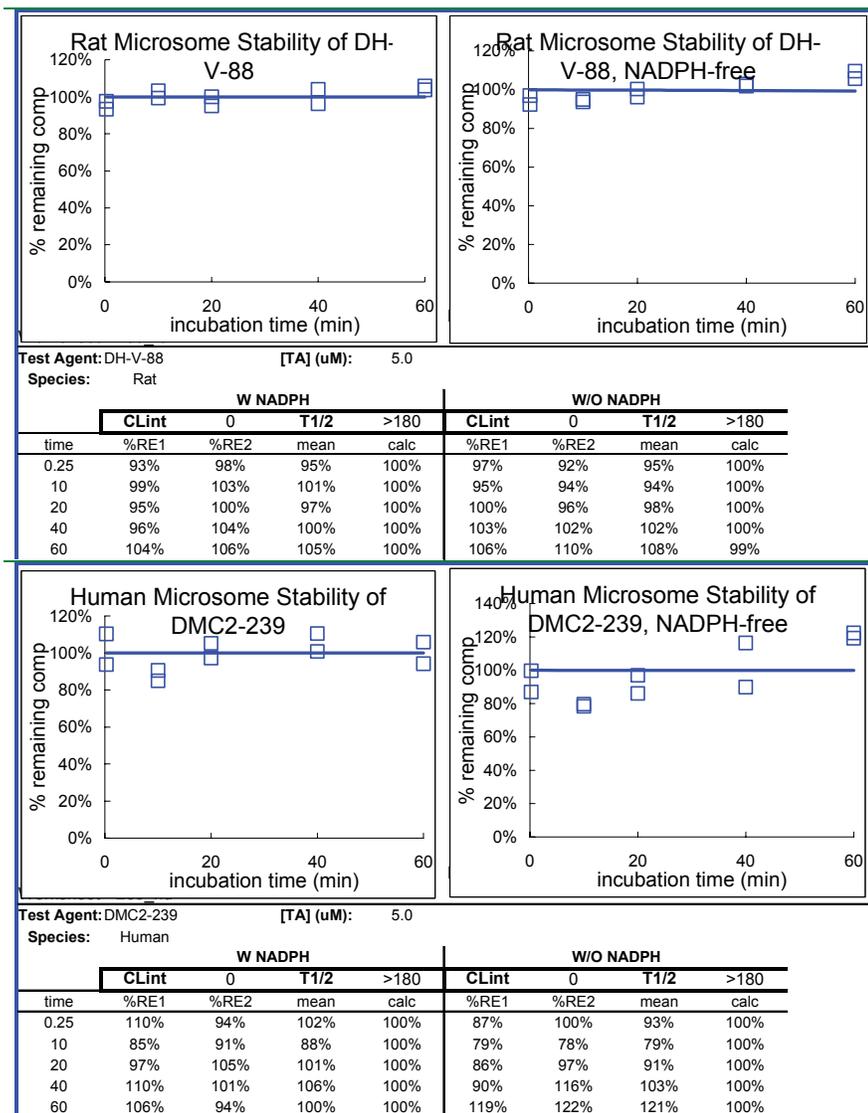
CONFIDENTIAL



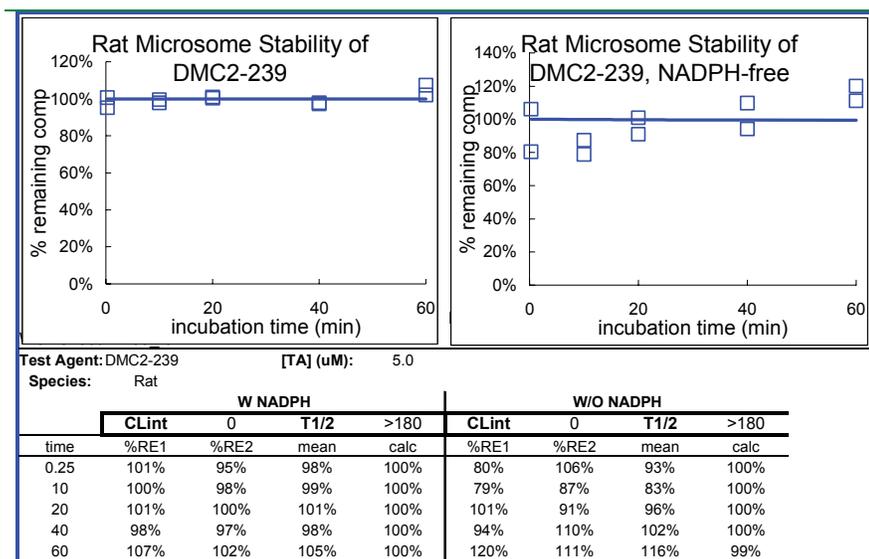
#### 4.3.5 Microsomal intrinsic clearance individual data



CONFIDENTIAL

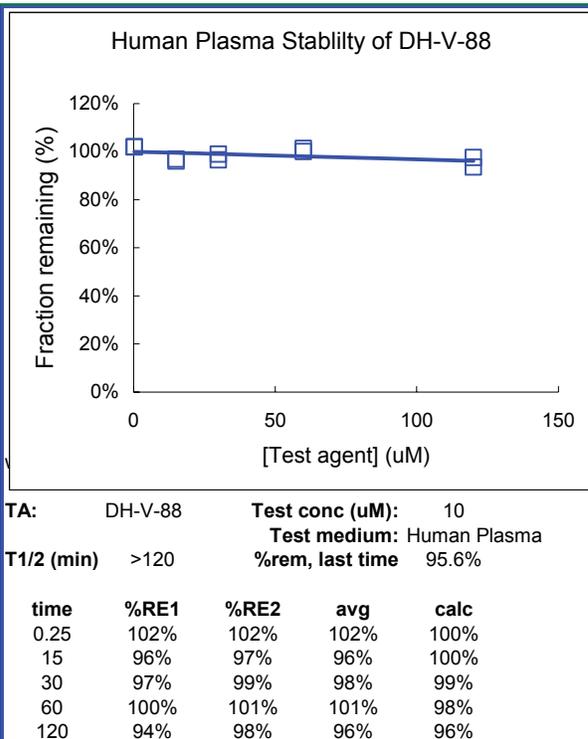


CONFIDENTIAL

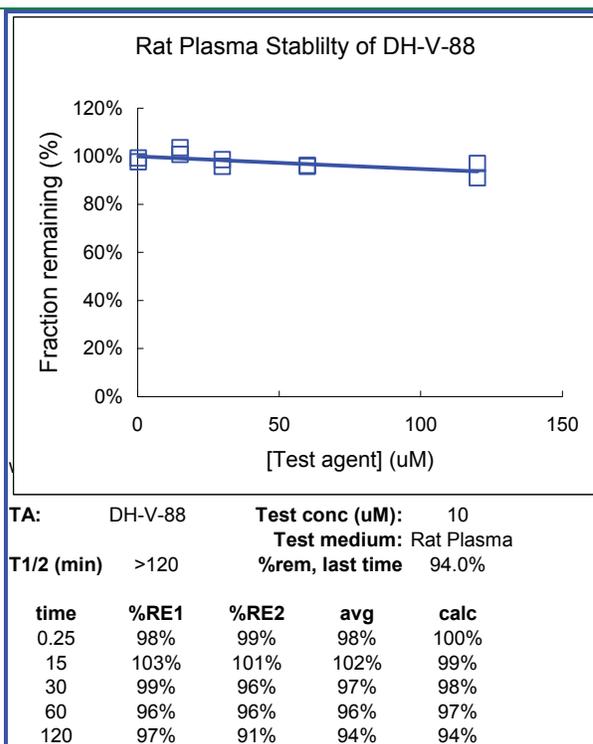


CONFIDENTIAL

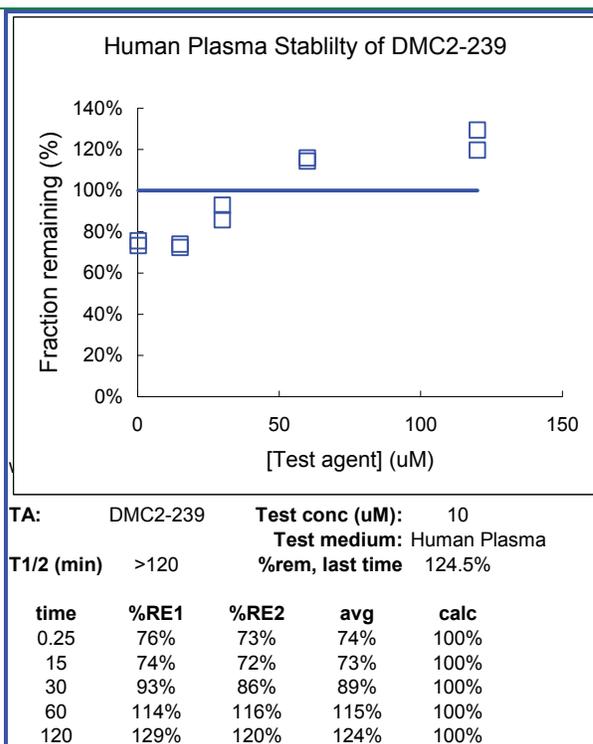
## 4.3.6 Plasma half-life individual data



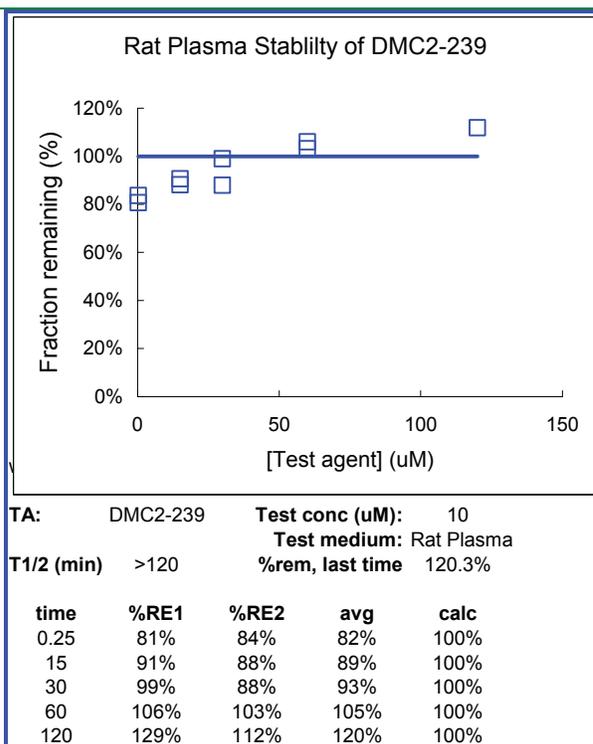
CONFIDENTIAL



CONFIDENTIAL



CONFIDENTIAL



#### 4.3.7 Plasma protein binding individual data

Client ID	test species	test conc (uM)	free fraction (%)			comment
			1 <sup>st</sup>	2 <sup>nd</sup>	mean	
DH-V-88	Human	10	0.00000%	0.0030%	0.0015%	
DH-V-88	Rat	10	0.00114%	0.0020%	0.0016%	
DMC2-239	Human	10	0.00000%	0.0000%	0.0000%	
DMC2-239	Rat	10	0.00000%	0.0079%	0.0040%	

<sup>a</sup>peak area(analyte) / peak area (internal standard)

## 5 References

Stewart, BH, *et al.* (1995) "Comparison of intestinal permeabilities determined in multiple *in vitro* and *in situ* models: Relationship to absorption in humans." *Pharm. Res.* 12:693.

Crespi, CL, Stresser, DM (2000) "Fluorometric screening for metabolism-based drug-drug interactions." *J. Pharmacol. Toxicol. Methods.* 44:325.

Houston, JB (1994) "Utility of *in vitro* drug metabolism data in predicting *in vivo* metabolic clearance." *Biochem. Pharmacol.* 47:1469.

Banker, MJ, *et al.* (2003) "Development and validation of a 96-well equilibrium dialysis apparatus for measuring plasma protein binding." *J. Pharm. Sci.* 92:967.

## **6 Storage and Retention of Records**

All documents generated in this study (raw data, the study plan, a copy of this report, etc.) will be stored for three years from the date of this document. Only authorized Apredica employees will have access to the archives.

The original final report will be provided to the sponsor and will be kept by the sponsor under its sole responsibility.

## 7 Appendices

### 7.1 Appendix A. Standard Apredica Methods

#### Caco-2 monolayer permeability

CaCo-2 cells grown in tissue culture flasks are trypsinized, suspended in medium, and the suspensions were applied to wells of a collagen-coated BioCoat Cell Environment in 24-well format (BD Biosciences) at 24,500 cells per well. The cells are allowed to grow and differentiate for three weeks, feeding at 2-day intervals.

For Apical to Basolateral (A->B) permeability, the test agent is added to the apical (A) side and amount of permeation is determined on the basolateral (B) side; for Basolateral to Apical (B->A) permeability, the test agent is added to the B side and the amount of permeation is determine on the A side. The A-side buffer contains 100  $\mu$ M Lucifer yellow dye, in Transport Buffer (1.98 g/L glucose in 10 mM HEPES, 1x Hank's Balanced Salt Solution) pH 6.5, and the B-side buffer is Transport Buffer, pH 7.4. CaCo-2 cells are incubated with these buffers for 2 h., and the receiver side buffer is removed for analysis by LC/MS/MS.

To verify the CaCo-2 cell monolayers are properly formed, aliquots of the cell buffers are analyzed by fluorescence to determine the transport of the impermeable dye Lucifer Yellow.

Data are expressed as permeability ( $P_{app}$ ): 
$$P_{app} = \frac{dQ/dt}{C_0 A}$$

where dQ/dt is the rate of permeation,  $C_0$  is the initial concentration of test agent, and A is the area of the monolayer.

In bidirectional permeability studies, the asymmetry index (AI) is also calculated:

$$AI = \frac{P_{app}(B \rightarrow A)}{P_{app}(A \rightarrow B)}$$

An AI > 1 indicated a potential substrate for PGP or other active transporters.

#### Cytotoxicity

HepG2 human hepatocellular carcinoma cells (originally obtained from ATCC, Manassas, VA) are seeded in 96-well plates at 10,000 cells per well, and grown for 24 hr in Eagle's Modified Essential Medium supplemented with 2 mM glutamine, nonessential amino acids, 2 mM pyruvate, 10% fetal bovine serum, 100 U/mL penicillin, and 100 mg/L streptomycin at 37°C in 5% CO<sub>2</sub>. NIH/3T3 mouse fibroblasts (originally obtained from ATCC) are seeded in 96-well plates at 10,000 cells per well, and grown for 24 hr in Dulbecco's Modified Essential Medium supplemented with 10% bovine calf serum, 100 U/mL penicillin, and 100 mg/L streptomycin at 37°C in 5% CO<sub>2</sub>. HaCaT human keratinocytes are seeded in 96-well plates at 10,000 cells per well, and grown for 24 hr in Dulbecco's Modified Essential Medium supplemented with 10% fetal bovine serum, 100 U/mL penicillin, and 100 mg/L streptomycin at 37°C in 5% CO<sub>2</sub>.

Test article is aseptically diluted in DMSO to 200x the highest concentration, then 100-fold in the growth medium, and serial dilutions are made in 1% DMSO in growth medium. At the start of the assay, the growth medium is removed from the plates and replaced with fresh medium, and an equal volume from each test agent dilution is added. Cells are incubated with test article for 48 h, and the wells are examined microscopically to look for abnormalities.

CONFIDENTIAL

For Neutral red staining, medium is removed, the cells are washed with PBS, and fresh medium containing 25 µg/mL neutral red (Sigma) is added. After four hours incubation, the cells are washed with PBS, and the cellular dye is solubilized with 1% acetic acid in 50% ethanol. Cellular neutral red is measured by its absorbance at 540 nm. Cytotoxicity is assessed by determining the IC<sub>50</sub> (the concentration that causes 50% reduction in uptake of neutral red after 48 hrs exposure to compound).

For MTT staining, 20 µL 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, 5 mg/mL in PBS, Sigma), is added to each well. After two hours incubation, the medium is removed, and the cellular dye is solubilized with DMSO. Cellular-converted MTT is measured by its absorbance at 540 nm. Cytotoxicity is assessed by determining the IC<sub>50</sub> (the concentration that causes 50% reduction in uptake of MTT after 48 hrs exposure to compound).

### Fluorescent cytochrome P450 IC<sub>50</sub> determination

Cytochrome P450 inhibition is measured using fluorogenic substrates. Test agents and substrates are dissolved in acetonitrile for this assay, as DMSO significantly inhibits some cytochrome P450s. Assays were performed at 37 °C using commercially available recombinant human cytochrome P450 expressed in insect cells. Enzyme concentrations and reactions times are optimized for each batch of enzyme to ensure a linear production of product over the course of the reaction. Percent remaining activity is calculated by comparing product formation of wells treated with test agent against wells treated with vehicle, after subtraction of background fluorescence. Percent inhibition is 100% - percent remaining activity. IC<sub>50</sub> is calculated using a four-point logistic curve model. The individual reaction conditions are summarized in the following Table.

Cytochrome	Substrate	Assay Buffer
1A2	5 µM CEC	Buffer B
2A6	3 µM coumarin	Buffer D
2B6	15 µM MFC	Buffer B
2C8	1 µM DBF	Buffer E
2C9	1 µM DBF	Buffer A
2C19	2 µM DBF	Buffer B
2D6	1.5 µM ANMC	Buffer C
2E1	100 µM MFC	Buffer B
3A4	50 µM BFC	Buffer B
3A4	1 µM DBF	Buffer B
3A5	50 µM BFC	Buffer B
<b>Aromatase</b>	0.5 µM DBF	Buffer B
Buffer A: 1.3 mM NADP, 3.3 mM glucose-6-phosphate, 0.4 U/mL glucose-6-phosphate dehydrogenase, 25 mM potassium phosphate, 3.3 mM MgCl <sub>2</sub> , pH 7.4.		
Buffer B: 1.3 mM NADP, 3.3 mM glucose-6-phosphate, 0.4 U/mL glucose-6-phosphate dehydrogenase, 100 mM potassium phosphate, 3.3 mM MgCl <sub>2</sub> , pH 7.4.		
Buffer C: 8.2 µM NADP, 0.41 mM glucose-6-phosphate, 0.4 U/mL glucose-6-phosphate dehydrogenase, 100 mM potassium phosphate, 0.41 mM MgCl <sub>2</sub> , pH 7.4.		
Buffer D: 0.066 mM NADP, 3.3 mM glucose-6-phosphate, 0.4 U/mL glucose-6-phosphate dehydrogenase, 100 mM Tris hydrochloride, 3.3 mM MgCl <sub>2</sub> , pH 7.5.		
Buffer E: 1.3 mM NADP, 3.3 mM glucose-6-phosphate, 0.4 U/mL glucose-6-phosphate dehydrogenase, 50 mM potassium phosphate, 3.3 mM MgCl <sub>2</sub> , pH 7.4.		
Abbreviations: ANMC: 3-[2-(N,N-Diethyl-N-methylammonium)ethyl]-7-methoxy-4-methylcoumarin; BFC: 7-Benzyloxy-4-(trifluoromethyl)coumarin; CEC: 3-Cyano-7-ethoxycoumarin; DBF: dibenzylfluorescein; MFC: 7-Methoxy-4-(trifluoromethyl)coumarin		

### hERG FastPatch

**Cell culture.** HEK293 cells were stably transfected with hERG cDNA. Stable

transfectants have been selected by coexpression with the G418-resistance gene incorporated into the expression plasmid. Selection pressure was maintained by including G418 in the culture medium. Cells were cultured in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/ F-12) supplemented with 10% fetal bovine serum, 100 U/mL penicillin G sodium, 100 µg/mL streptomycin sulfate and 500 µg/mL G418. Before testing, cells in culture dishes were washed twice with Hank's Balanced Salt Solution, treated with trypsin and re-suspended in the culture media ( $1-1.5 \times 10^6$  cells in 20 mL). Cells in suspension were allowed to recover for 1-3 hours in a tissue culture incubator set at 37°C in a humidified 95% air, 5% CO<sub>2</sub> atmosphere. Immediately before use in the PatchXpress® system, the cells were washed in HB-PS to remove the culture medium and re-suspended in 150 µL of HB-PS.

**Test Method.** All experiments were performed at ambient temperature. Each cell acted as its own control.

**Test Article Treatment Groups.** Two concentrations were applied at five (5) minute intervals via disposable polyethylene micropipette tips to cells expressing hERG ( $n \geq 2$ , where  $n$  = the number cells/concentration). Each solution exchange, performed in quadruplicate, consisted of aspiration and replacement of 45 µL of the total 50 µL volume of the extracellular well of the Sealchip16. Duration of exposure to each test article concentration was five (5) minutes.

**Test Article Application Schedule.**

Solution	Procedure	Exposure time
Vehicle control	four 45 µL exchanges	10 min
Test article concentration 1	four 45 µL exchanges	5 min
Test article concentration 2	four 45 µL exchanges	5 min

**Positive Control Treatment Group.** Vehicle was applied to cells expressing hERG ( $n \geq 2$ , where  $n$  = the number cells), for a 10-minute exposure interval. Each solution exchange, performed in quadruplicate, consisted of aspiration and replacement of 45 µL of the total 50 µL volume of the extracellular well of the Sealchip16. After vehicle application, the positive control was applied in the same manner, to verify sensitivity to hERG blockade.

**Automated Patch Clamp Electrophysiological Procedures.** Intracellular solution for whole cell recordings consisted of (composition in mM): potassium aspartate, 130; MgCl<sub>2</sub>, 5; EGTA, 5; ATP, 4; HEPES, 10; pH adjusted to 7.2 with KOH. This solution was prepared in batches, aliquoted, stored frozen, and a fresh aliquot thawed each day. In preparation for a recording session, intracellular solution was loaded into the intracellular compartments of the Sealchip16 planar electrode. Cell suspension was pipetted into the extracellular compartments of the Sealchip16 planar electrode. After establishment of a whole-cell configuration, membrane currents were recorded using dual-channel patch clamp amplifiers in the PatchXpress® system. Before digitization, the current records were low-pass filtered at one-fifth of the sampling frequency.

**Voltage-Clamp Procedures.** Onset and block of hERG current was measured using a stimulus voltage pattern (Figure 1, lower panel) consisting of a 500 ms prepulse to -40 mV (leakage subtraction), a 2-second activating pulse to +40 mV, followed by a 2-second test pulse to -40 mV. The pulse pattern was repeated continuously at 10 s intervals, from a holding potential of -80 mV. Peak tail current (Figure 1, upper panel) was measured during the -40 mV test pulse. Leakage current was calculated from the current amplitude evoked by the prepulse and subtracted from the total membrane current record.

**Microsomal intrinsic clearance**

The test agent is incubated in duplicate with microsomes at 37 °C. The reaction contains microsomal protein in 100 mM potassium phosphate, 2 mM NADPH, 3 mM MgCl<sub>2</sub>, pH 7.4. A control is run for each test agent omitting NADPH to detect NADPH-free degradation. The indicated times, an aliquot is removed from each experimental and

CONFIDENTIAL

control reaction and mixed with an equal volume of ice-cold Stop Solution (0.3% acetic acid in acetonitrile containing haloperidol, diclofenac, or other internal standard). Stopped reactions are incubated at least ten minutes at -20 °C, and an additional volume of water is added. The samples are centrifuged to remove precipitated protein, and the supernatants are analyzed by LC/MS/MS to quantitate the remaining parent. Data are converted to % remaining by dividing by the time zero concentration value. Data are fit to a first-order decay model to determine half-life. Intrinsic clearance is calculated from the half-life and the protein concentrations:  $CL_{int} = \ln(2) / (T_{1/2} [\text{microsomal protein}])$ .

#### **Plasma protein binding**

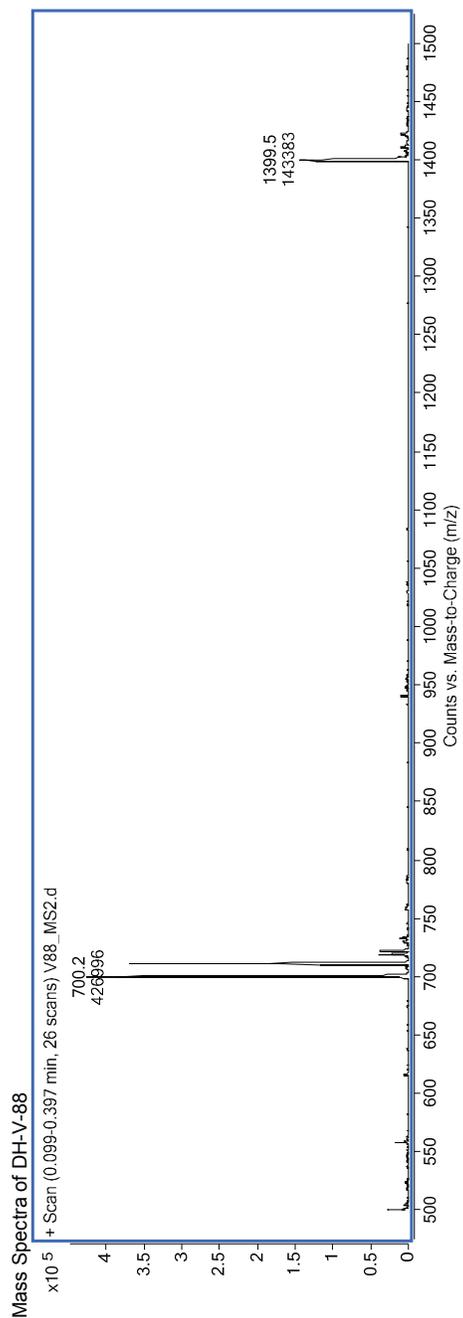
Test agent is added to plasma. This mixture is dialyzed in a RED Device (Pierce) per the manufacturers' instructions along against PBS and incubated in a rocker. After the end of the incubation, aliquots from both plasma and PBS sides are collected, an equal amount of PBS is added to the plasma sample, and an equal volume of plasma is added to the PBS sample. Methanol (three volumes) with haloperidol IS are added to precipitate the proteins and release the agents. After centrifugation, the supernatant was transferred to a new plate and analyzed by LC/MS/MS.

#### **Plasma half-life**

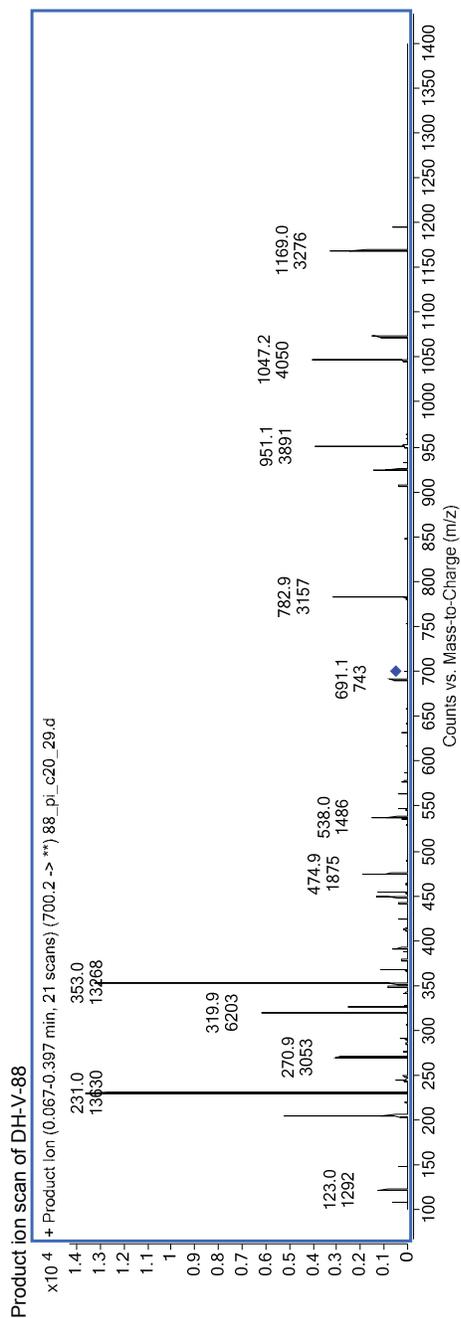
The test agent is incubated in duplicate with plasma at 37 °C. The reaction contains plasma and 2% DMSO. At the indicated times, an aliquot is removed from each experimental reaction and mixed with three volumes of ice-cold Stop Solution (methanol containing propranolol, diclofenac, or other internal standard). Stopped reactions are incubated at least ten minutes at -20 °C. The samples are centrifuged to remove precipitated protein, and the supernatants are analyzed by LC/MS/MS to quantitate the remaining parent. Data are converted to % remaining by dividing by the time zero concentration value. Data are fit to a first-order decay model to determine half-life.

CONFIDENTIAL

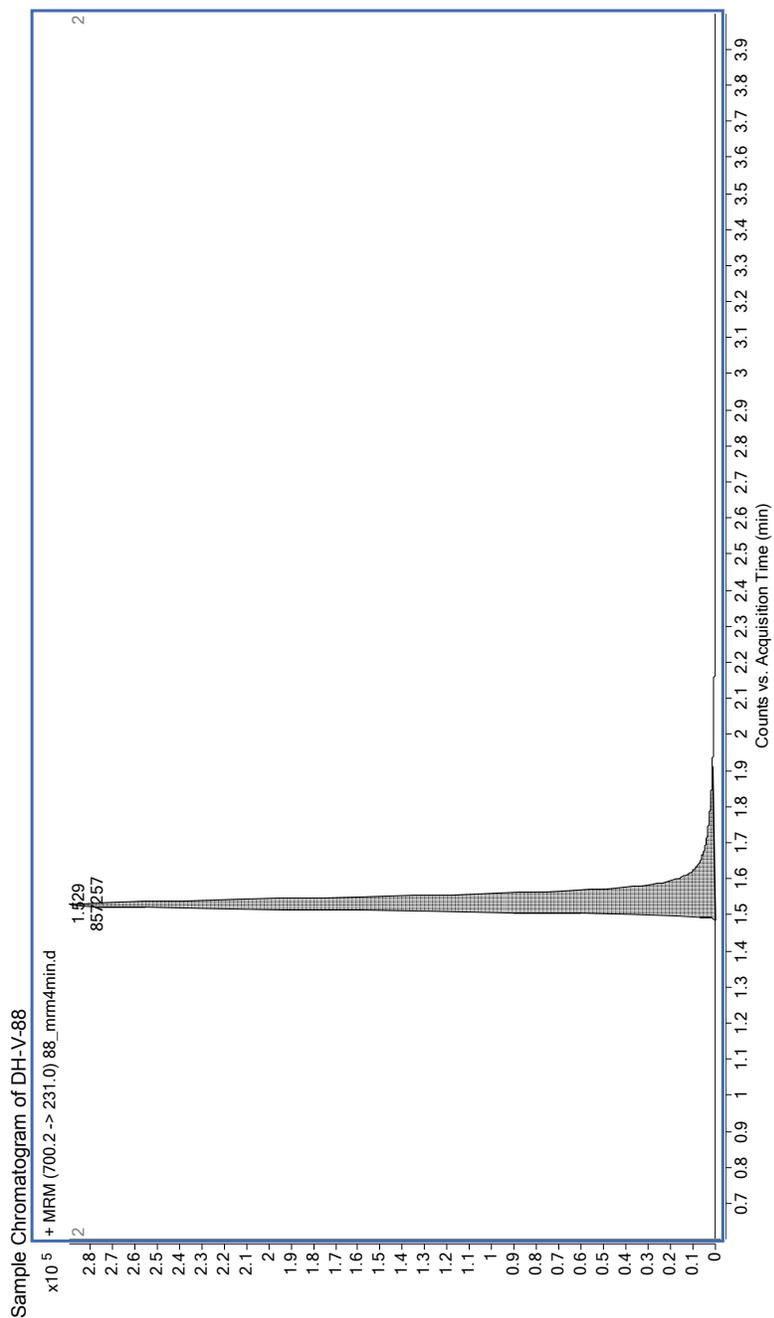
### 7.2 Appendix B. Sample Spectra and Chromatograms of the Test Agents



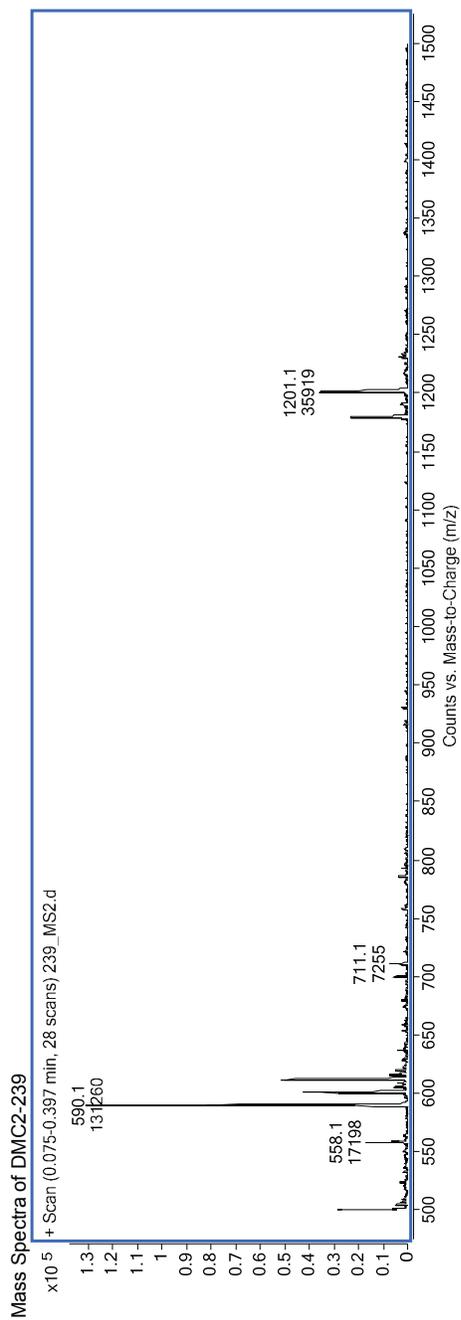
CONFIDENTIAL



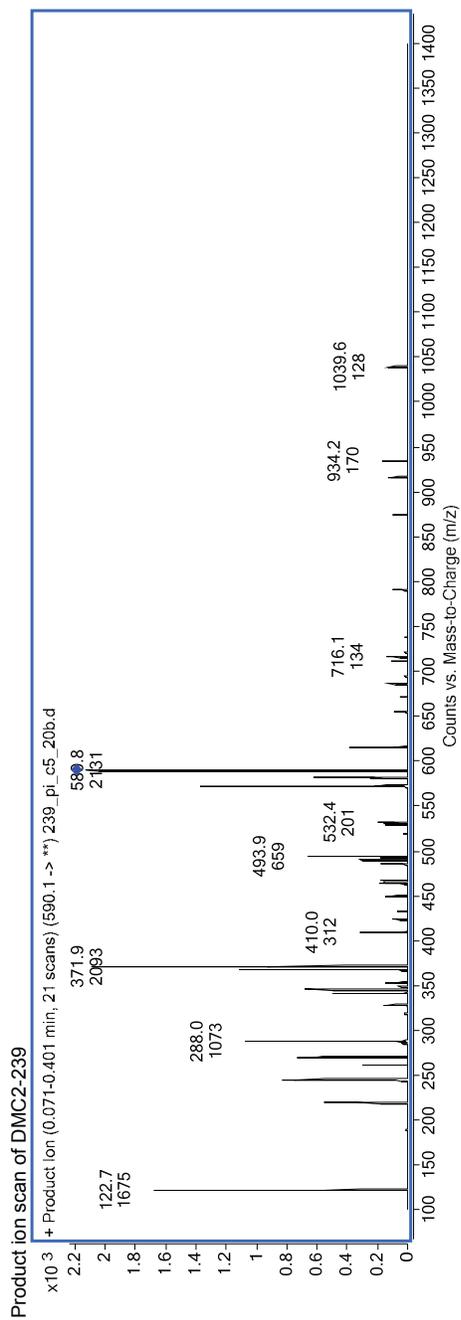
CONFIDENTIAL



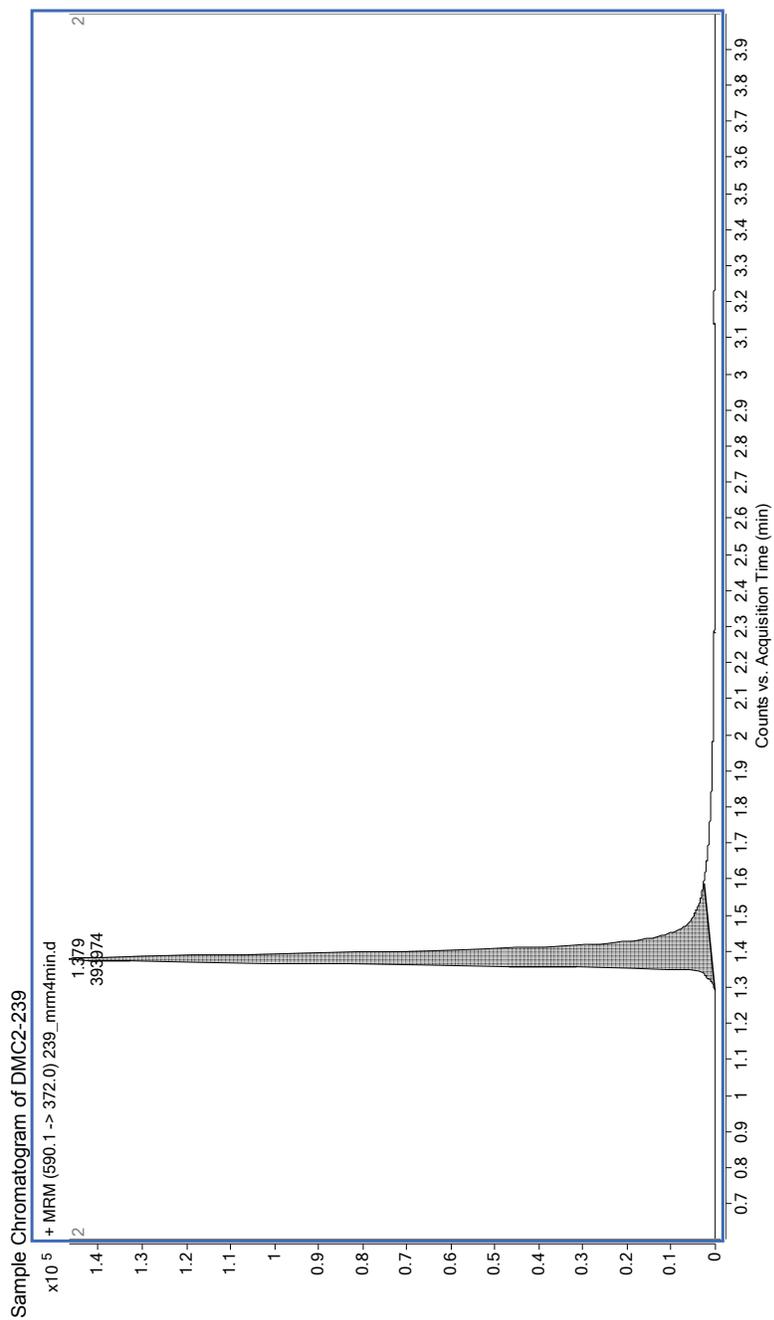
CONFIDENTIAL



CONFIDENTIAL



CONFIDENTIAL



## Appendix C: Programmable Oligomers for Minor Groove DNA

### Recognition

*The text of this chapter was taken in part from a manuscript coauthored with Raymond M. Doss, Michael M. Marques, Shane Foister, and Peter B. Dervan\* (Caltech)*

(Doss, R.M., Marques, M.A., Foister, S., Chenoweth, D.M., and Dervan, P. B. *J. Am. Chem. Soc.*, **2006**, *128*, 9074-9079.)

**Abstract**

The four Watson-Crick base pairs of DNA can be distinguished in the minor groove by pairing side-by-side three five-membered aromatic carboxamides, imidazole (Im), pyrrole (Py), and hydroxypyrrole (Hp), four different ways. On the basis of the paradigm of unsymmetrical paired edges of aromatic rings for minor groove recognition, a second generation set of heterocycle pairs, imidazopyridine/pyrrole (Ip/Py) and hydroxybenzimidazole/pyrrole (Hz/Py), revealed that recognition elements not based on analogues of distamycin could be realized. A new set of end-cap heterocycle dimers, oxazole-hydroxybenzimidazole (No-Hz) and chlorothiophene-hydroxybenzimidazole (Ct-Hz), paired with Py-Py are shown to bind contiguous base pairs of DNA in the minor groove, specifically 5'-GT-3' and 5'-TT-3', with high affinity and selectivity. Utilizing this technology, we have developed a new class of oligomers for sequence-specific DNA minor groove recognition no longer based on the *N*-methyl pyrrole carboxamides of distamycin.

## C.1 Introduction

Aberrant gene expression is the cause of many diseases, and the ability to reprogram transcriptional pathways using cell-permeable small molecules may, one day, have an impact on human medicine.<sup>1</sup> DNA-binding polyamides, which are based on the architecture of the natural products netropsin and distamycin A,<sup>2,3a,b</sup> are capable of distinguishing all four Watson-Crick base pairs in the DNA minor groove and have been the subject of intense study along with many other classes of minor groove binders.<sup>3c-f,4,5</sup> Sequence-specific recognition of the minor groove of DNA by polyamides arises from the pairing of three different antiparallel five-membered heterocyclic amino acids, pyrrole (Py), imidazole (Im), and hydroxypyrrole (Hp).<sup>4,5</sup> The direct read out, or information face, on the inside of the crescent-shaped polyamide may be programmed by the incremental change of atoms on the corners of the ring pairs presented to the DNA minor groove floor. Stabilizing and, importantly, destabilizing interactions with the different edges of the four Watson-Crick bases are modulated by shape complementarity and specific hydrogen bonds.<sup>4,6,7l</sup> For example, the Im/Py pair distinguishes G•C from C•G, T•A, and A•T. Im presents a lone pair of electrons to the DNA minor groove and can accept a hydrogen bond from the exocyclic amine of guanine.<sup>5</sup> Additionally, the Hp/Py pair distinguishes T•A from A•T, G•C, and C•G.<sup>4,6</sup> Hp projects an exocyclic OH group toward the minor groove floor that is sterically accommodated in the cleft of the T•A base pair, preferring to lie over T, not A.<sup>5</sup> These pairing rules have proven useful for programmed recognition of a broad repertoire of DNA sequences; however, the hydroxypyrrole ring system has proven to be unstable over time and in the presence of acid, further prompting our search for new T•A/A•T recognition elements. In addition, sequence-dependent changes in the microstructure of DNA (intrinsic minor groove width, minor groove flexibility, and inherent DNA curvature)<sup>7a-k</sup> combined with structural and conformational changes among polyamides make the targeting of certain sequences less than optimal, leading us to explore whether other novel heterocyclic recognition elements could be discovered for use in DNA groove recognition within the unsymmetrical pairing paradigm.<sup>7l,8-10</sup> Furthermore, from a medicinal chemistry point of view, a broader tool kit of sequence-specific recognition elements for DNA beyond polyamides would be useful as our artificial transcription factor program moves from cell culture<sup>11</sup> to small animal studies.

We recently reported that the benzimidazole ring can be an effective platform for the development of modular paired recognition elements for the minor groove of DNA.<sup>9,10</sup> The benzimidazole 6-5 bicyclic ring structure, though having slightly different curvature from the classic five-membered pyrrole-carboxamides, presents an “inside edge” with a similar atomic readout to

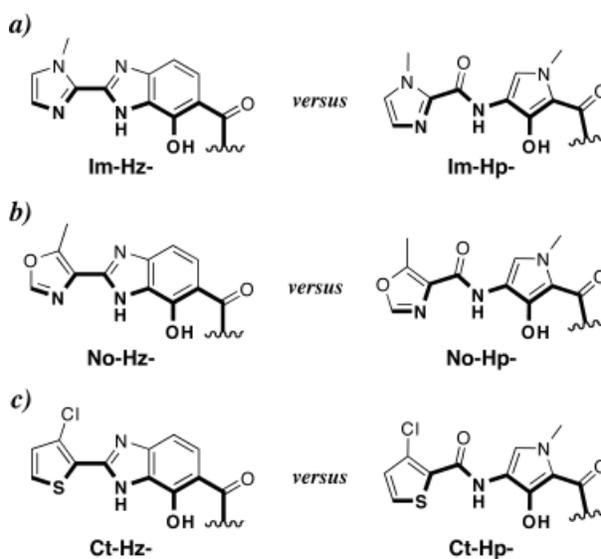
the DNA minor groove floor, effectively mimicking Py, Im, and Hp. We demonstrated that the imidazopyridine/pyrrole pair Ip/Py distinguishes G•C from C•G and the hydroxybenzimidazole/pyrrole pair Hz/Py distinguishes T•A from A•T, providing a solution to the unanticipated hydroxypyrrole instability limitation.<sup>9,10</sup> The question arises whether this second generation solution to DNA recognition can be elaborated further, *deleting incrementally almost all carboxamide linkages in the backbone of the hairpin motif*.<sup>12</sup>

We report here a new set of heterocycle dimer pairs,<sup>12</sup> which represents a step from single base-pair recognition toward a two letter approach to molecular recognition of the minor groove of DNA (Figure C.1). We move from single letters to syllables. New heterocycles were designed by combining the T-specific hydroxybenzimidazole (Hz) with oxazole (No) rings and chlorothiophene (Ct) caps<sup>8</sup> to afford the recognition elements No-Hz and Ct-Hz, respectively (Figure C.1). Quantitative DNase I footprinting titrations were used to determine DNA binding affinities of hairpin oligomers containing the No-Hz and Ct-Hz dimers paired with Py-Py dimer for each of the four Watson-Crick bases (Figure C.2). When positioned at the termini of hairpin polyamides, the No-Hz/Py-Py and Ct-Hz/Py-Py dimer pairs are found to target 5'-GT-3' and 5'-TT-3' sequences, respectively, with high affinity and good specificity. With the development of dimer pairs capable of recognizing a 5'-GT-3' sequence of DNA,

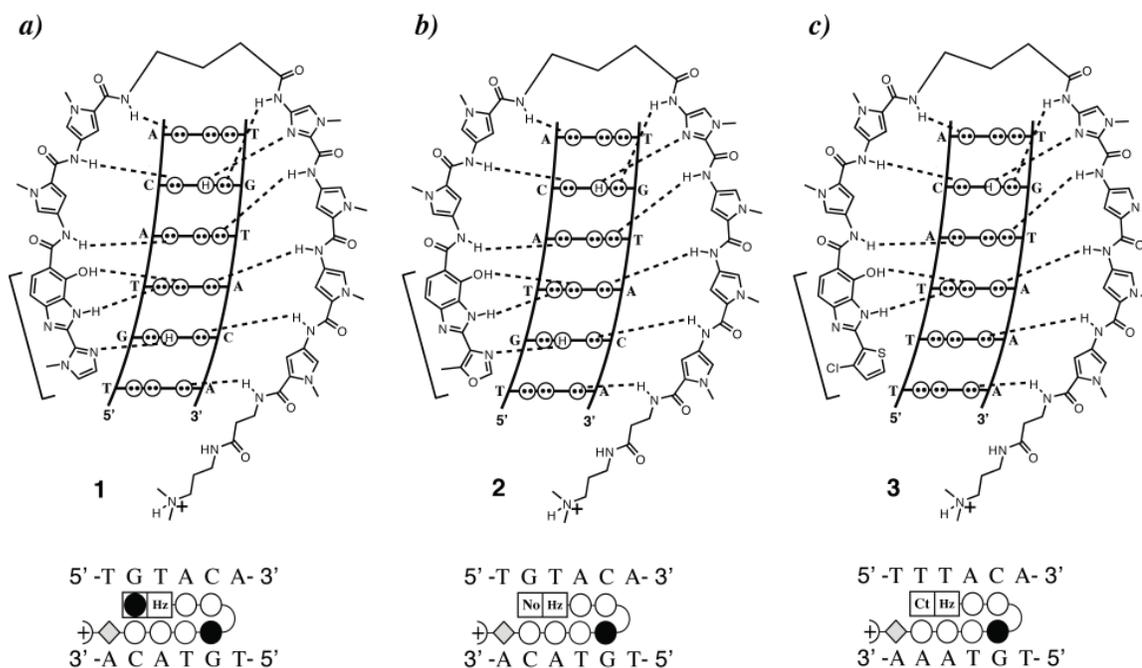
we could address the question whether a hairpin oligomer comprised of four dimer units will bind the site 5'-GTAC-3', a sequence formally containing all four Watson-Crick base pairs. Such a molecule represents our first programmable oligomer, which demonstrates excellent DNA binding properties without containing a single pyrrole or imidazole-carboxamide based on the natural product distamycin, a design benchmark for biomimetic chemistry and the field of DNA recognition.

## C.2 Experimental

### C.2.1 Polyamide Synthesis



**Figure C.1** Structures of dimers. a) imidazole-hydroxybenzimidazole (Im-Hz), b) oxazole-hydroxybenzimidazole (No-Hz), and c) chlorothiophenehydroxybenzimidazole (Ct-Hz) dimer caps in comparison with their respective five membered ring systems. Hydrogen-bonding surfaces to the DNA minor-groove floor are bolded.



**Figure C.2** Postulated hydrogen-bonding models for the 1:1 polyamide-DNA complexes with their matched sequence and their ball-and-stick representations. a) Im-Hz-Py-Py- $\gamma$ -Im-Py-Py-Py- $\beta$ -Dp (**1**), b) No-Hz-Py-Py- $\gamma$ -Im-Py-Py-Py- $\beta$ -Dp (**2**), and c) Ct-Hz-Py-Py- $\gamma$ -Im-Py-Py-Py- $\beta$ -Dp (**3**).

Hairpin polyamides were synthesized manually from Boc- $\beta$ -PAM resin in a stepwise fashion using Boc-protected monomeric and dimeric amino acids according to established solid-phase protocols.<sup>13</sup> Base Resin 1 (**BR1**) ( $\text{H}_2\text{N-Py-Py-}\gamma\text{-Im-Py-Py-Py-}\beta\text{-Pam}$ ) was synthesized in gram quantities using the following amino acid building blocks: Boc-Py-OBt (**4**), Boc-Im-OH (**5**), and Boc- $\gamma$ -OH (**6**) (Scheme C.1). The base resins were then split into smaller batches for coupling to the final dimeric caps. Boc-protected amino acid monomers for Boc-Py-OBt (**4**) and Boc-Im-OH (**5**) were synthesized according to previously reported procedures.<sup>8,13,14</sup> Dimeric cap synthesis for No-HzOMe-OH (**7**) and Ct-HzOMe-OH (**8**) are detailed in Section C.7 Supplemental Information. Couplings were achieved using preactivated monomers (Boc-Py-OBt) or HBTU activation in a DIEA and DMF mixture. Coupling times ran from 3 to 24 h at 25-40°C. Deprotection of the growing polyamide was accomplished using 80% TFA/DCM. Polyamides were cleaved from the resin by treatment with neat 3-(dimethylamino)-1-propylamine (Dp) at 80°C for 2 h and purified by preparatory reverse phase HPLC. Deprotection of the methoxy-protected polyamides was done using a mix of thiophenoxide in DMF at 80°C to provide the free hydroxy derivatives after a second HPLC purification: Im-Hz-Py-Py- $\gamma$ -Im-Py-Py-Py- $\beta$ -Dp (**1**), No-Hz-Py-Py- $\gamma$ -Im-Py-Py-Py- $\beta$ -Dp (**2**), Ct-Hz-Py-Py- $\gamma$ -Im-Py-Py-Py- $\beta$ -Dp (**3**). (See Section C.7 Supplemental Information for full experimental details.)

### C.3 Results

#### C.3.1 DNA Affinity and Sequence Specificity of Dimer Caps

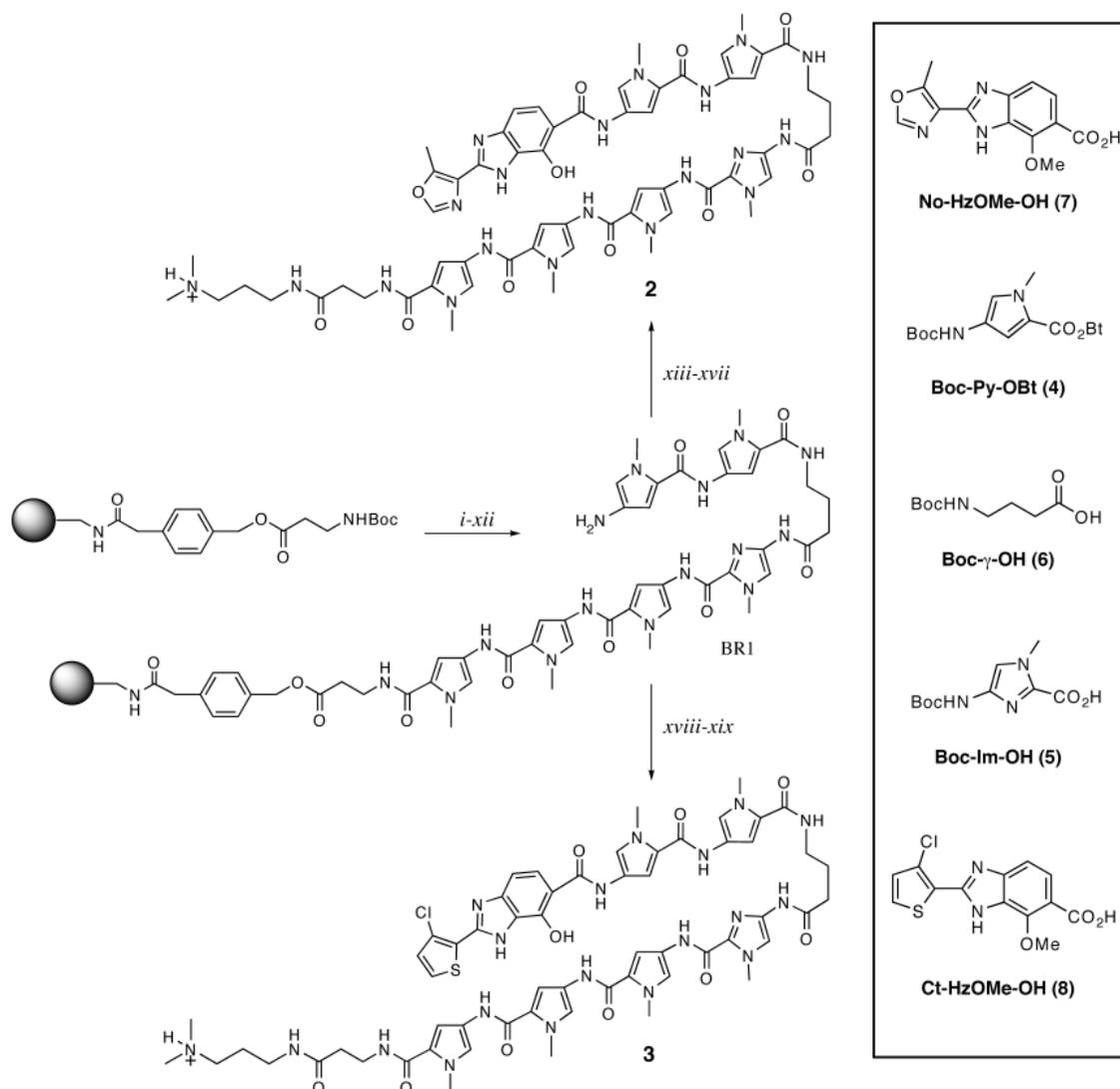
Quantitative DNase I footprinting titrations were carried out for polyamides **1-3**. All polyamides were footprinted on the 285-base-pair PCR product of plasmid pCW15. In all cases, the DNA-sequence specificity at the cap position (in bold) was determined by varying a single DNA base pair within the sequence, 5'-**T**XTACA-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes. The variable base-pair position was designed to be adjacent to the Hz ring, which has been shown to specify for T when paired across from Py, so as to be able to determine the binding properties of each compound to the following two base-pair sequences: AT, TT, GT, and CT.

The sequence specificity of the Im-Hz and Ct-Hz dimers for 5'-**T**XTACA-3' were evaluated in polyamides **1** and **3**, respectively. As expected, polyamide **3** bound its designed match site 5'-**T**TTACA-3' ( $K_a = 2.4 \times 10^9 \text{ M}^{-1}$ ) (Figure C.3, Table C.1) with both the Ct and the Hz halves of the dimer preferring to rest over the less bulky T in the asymmetric cleft of a T•A base pair. Placing the Ct ring adjacent to the Hz resulted in a 10-fold specificity for T > A using the Ct-Hz system. Polyamide **1**, which contains the Im-Hz dimer, did not bind its designed match site 5'-**T**GTACA-3' with any appreciable level of specificity exhibiting affinities of  $K_a = 1.6 \times 10^8$  and  $4.0 \times 10^8 \text{ M}^{-1}$  for the GT and AT sites, respectively.

Oxazole cap (polyamide **2**) was incorporated into the dimer cap system, and the affinity for its designed match site, 5'-**T**GTACA-3', was examined. Polyamide **2** successfully targeted its designed match site with an appreciable level of specificity (25-fold) and a match site affinity of  $K_a = 6.8 \times 10^9 \text{ M}^{-1}$  (Figure C.3, Table C.1). With the development of the chlorothiophene and oxazole dimer caps, the range of targetable sequences by polyamides has been expanded (Table C.1).

#### C.3.2 Design of a Programmable Oligomer for 5'-GTAC-3'

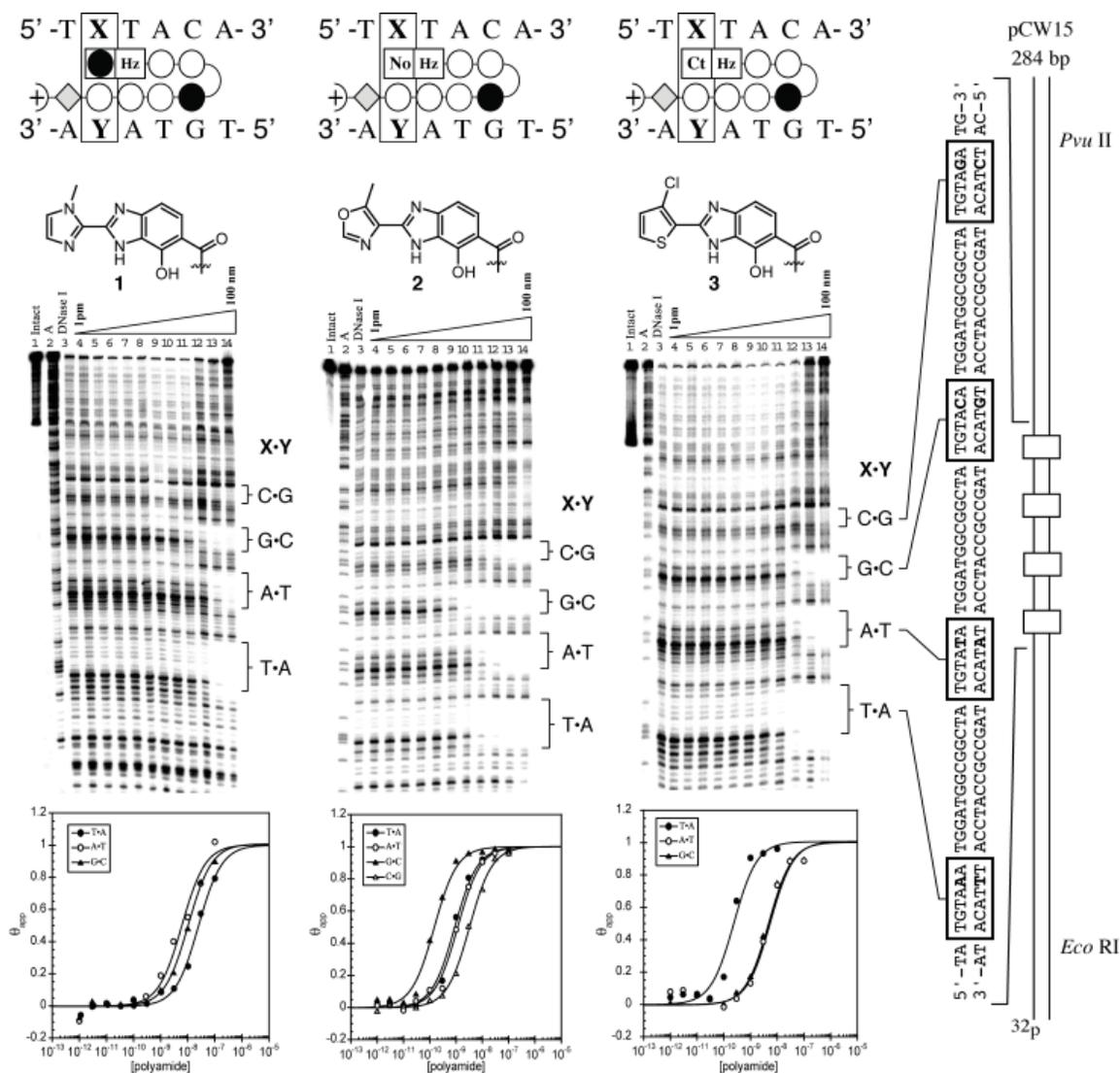
The synthesis of oligomer **9** containing four dimer units was achieved via the stepwise addition of Boc-amino acid dimers in the same manner as previously described polyamide syntheses. This "third generation" oligomer's binding properties were assessed in the same context<sup>16</sup> as previously reported for first and second generation hairpin polyamides targeting the sequence, 5'-GTAC-3', containing the four Watson-Crick base pairs.<sup>4,10</sup> Footprinting of the oligomer on the previously characterized plasmid DEH10 showed a binding affinity of  $K_a = 2.3 \times 10^{10} \text{ M}^{-1}$  for the



**Scheme C.1** Representative solid-phase synthesis of polyamide **2** and **3** along with a table of the amino acid building blocks used. Reaction conditions: (i) 80% TFA/DCM; (ii) Boc-Py-OBt, DIEA, DMF; (iii) Ac<sub>2</sub>O, DIEA, DMF; (iv) repeat i-iii 2; (v) 80% TFA/DCM; (vi) Boc-Im-OH, HBTU, DIEA, DMF; (vii) Ac<sub>2</sub>O, DIEA, DMF; (viii) 80% TFA/DCM; (ix) Boc- $\gamma$ -OH, HBTU, DIEA, DMF; (x) Ac<sub>2</sub>O, DIEA, DMF; (xi) repeat i-iii x 2; (xii) 80% TFA/DCM; (xiii) No-HzOMe-OH, HBTU, DIEA, DMF; (xiv) 3-(dimethylamino)-1-propylamine (Dp), 80 °C, 2 h; (xv) preparative HPLC; (xvi) thiophenol, NaH, DMF; (xvii) preparative HPLC; (xviii) Ct-HzOMe-OH, HBTU, DIEA, DMF; (xix) same as steps xiv-xvii.

match site 5'-GTAC-3' and affinities of  $K_a = 3.5 \times 10^9$  and  $9.8 \times 10^8 \text{ M}^{-1}$  for the mismatch sites 5'-GAAC-3' and 5'-GATC-3', respectively (Figure C.4). Such a result demonstrates that a compound consisting exclusively of 6-5 fused ring systems and minimal carboxamide linkages is able to maintain good levels of specificity and excellent binding affinity.

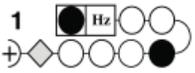
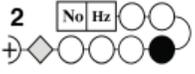
#### C.4 Discussion



**Figure C.3** Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **1**, **2**, and **3**, respectively, on the 278 bp, 5'-end-labeled PCR product of plasmid CW15: (lane 1) intact DNA; (lane 2) A reaction; (lane 3) DNase I standard; (lanes 4-14) 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (top) Chemical structure of the pairing of interest; (bottom) binding isotherms for the four designed sites.  $\theta_{\text{norm}}$  values were obtained according to published methods.<sup>15</sup> A binding model for the hairpin motif is shown centered at the top as a ball-and-stick model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and nonfilled circles, respectively;  $\beta$ -alanine is shown as a diamond; the  $\gamma$ -aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

Recent advances in hairpin polyamide designs have traditionally focused on developing new modes of single base-pair recognition by heterocyclic ring pairings. Previous studies, however, have highlighted the fact that the microstructure of DNA depends on the sequence in question.<sup>17</sup> In addition, structural and conformational changes among polyamides are thought to have an impact

**Table C.1** Affinities of X/Py ring pairs proximal to a hydroxybenzimidazole bicycle  $K_a$  ( $M^{-1}$ ).<sup>a,b</sup>

Polyamide	5'-tATACa-3'	5'-tTTACa-3'	5'-tGTACa-3'	5'-tCTACa-3'
<b>1</b> 	9.7( $\pm$ 0.7) $\times 10^7$	4.5( $\pm$ 0.6) $\times 10^8$	1.7( $\pm$ 0.4) $\times 10^8$	$\leq 1.0 \times 10^7$
<b>2</b> 	8.6( $\pm$ 0.3) $\times 10^8$	9.5( $\pm$ 0.3) $\times 10^8$	6.8( $\pm$ 0.4) $\times 10^9$	2.7( $\pm$ 0.5) $\times 10^8$
<b>3</b> 	2.1( $\pm$ 0.3) $\times 10^8$	2.4( $\pm$ 0.2) $\times 10^9$	2.6( $\pm$ 0.4) $\times 10^8$	$\leq 1.0 \times 10^7$

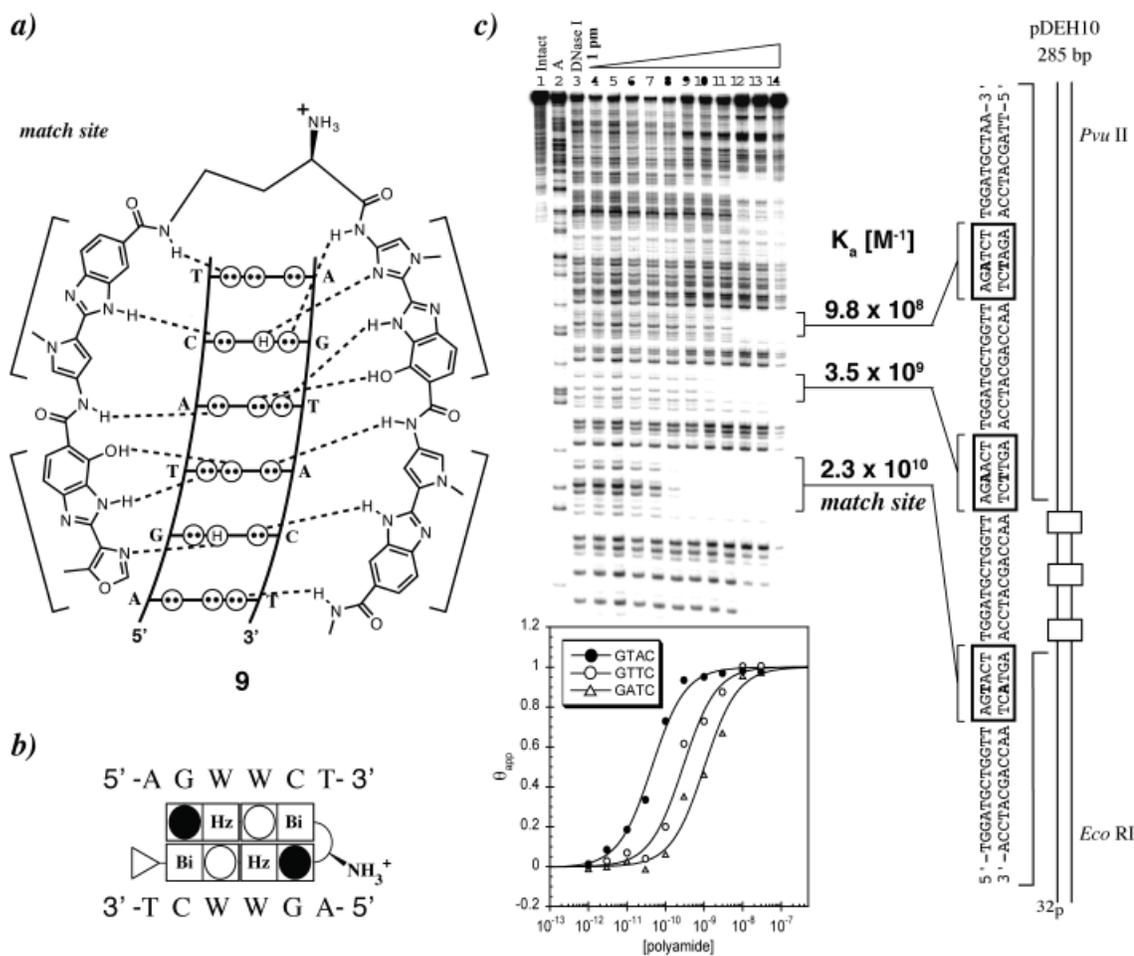
a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses.

b) Assays were performed at 22°C in a buffer of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5mM CaCl<sub>2</sub> at pH 7.0.

on DNA affinity and sequence specificity. Thus, we have taken a more global view of molecular recognition, where our efforts have expanded from designing modules that distinguish the four DNA base-pairs (i.e., pairing rules) to designing those that target short, discrete DNA sequences.

Upon incorporation and evaluation of the G specific Im ring into the Im-Hz dimer cap (polyamide **1**), we were surprised to find that it failed to demonstrate any preference for its designed site in addition to displaying a significantly decreased affinity. The shortcomings of the Im-Hz dimer prompted a search for a ring system that was capable of specifying for G > C within the X-Hz context. The oxazole (No) cap (Figure C.1) was considered because of its structural resemblance to Im (both rings present a nitrogen atom capable of hydrogen bonding to the minor groove). When the No-Hz dimer was incorporated into polyamide **2**, it was found to be specific for its designed sequence of 5'-TGTACA-3' with a 25-fold preference for G > C and an affinity of  $K_a = 6.8 \cdot 10^9 M^{-1}$  at its match site (Figure C.3). The No-Hz dimer presents the same functionality to the minor groove as the Im-Hz dimer but with an enhanced ability to target a 5'-GT-3' site, which could be due to a combination oxazole lone-pair basicity and differential solvation/desolvation effects between oxazole and imidazole.

The Ct-Hz dimer cap represents our first effort to target a short sequence of DNA using sequence-inspired recognition elements. Studies have shown that Hz exhibits specificity for T > A at the N-1 positions relative to the polyamide N-terminus and that Ct polyamides exhibited specificity for T > A at the cap position with excellent polyamide affinities. We hoped that a hybrid dimer would impart specificity for the TT sequence while maintaining a biologically relevant affinity. Polyamide **3** bound its designed match site 5'-TTTACA-3' with an affinity of  $K_a = 2.4 \times$



**Figure C.4** a) Postulated hydrogen-bonding model and structure of oligomer **9**. b) Ball and stick representation of **9** and the 6-base-pair binding site with variable region (W = A or T) shown. c) Quantitative DNase I footprint titration experiment on the 5'- $^{32}\text{P}$ -labeled PCR product shown with an illustration and complete sequence of the 285 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pDEH10. Binding affinities are shown next to their respective binding sites and the match site is designated.

$10^9$  M $^{-1}$  and a specificity of 10-fold for T•A over A•T. This result is attributed to the fact that both the sulfur and hydroxyl groups prefer to lie over the less-bulky thymine base in a T•A base pair and that the  $-\text{OH}$  of the Hz ring is able to form an energetically favorable hydrogen bond with the O(2) carbonyl of thymine.<sup>10</sup> Combined, these attributes make this dimer the preferred solution for targeting consecutive thymine residues.

As a first step in the design of programmable oligomers devoid of directly linked pyrrole- or imidazole-carboxamides, we incorporated the new No-Hz dimer into a hairpin structure, oligomer **9**, consisting only of 6-5 fused ring systems (Figure C.2). The oligomer was designed to target the site 5'-GTAC-3' and is a third generation molecule from our previously reported hairpin polyamides,

which were shown to code for the four Watson-Crick base pairs in a sequence specific manner.<sup>16</sup> To evaluate the impact of removing four carboxamide linkages and moving to a system consisting of only 6-5 fused recognition elements, binding properties were evaluated using quantitative DNase I footprinting titrations. Oligomer **9** was found to bind its match site with an impressive affinity of  $K_a = 2.3 \times 10^{10} \text{ M}^{-1}$  while discriminating against its mismatch sites of 5'-GAAC-3' and 5'-GATC-3' with specificities ( $K_a \text{ match}/K_a \text{ mismatch}$ ) of ~7- and ~23-fold, respectively (Figure C.4). The four 6-5 fused rings of oligomer **9** present an "inside edge" with complimentary shape to the minor groove floor as the carboxamide linkages of traditional Py-Im-Hp polyamides. The complementary bumps and holes fit together between the oligomer and the DNA surface, which is the key to specificity. The dimer recognition elements in oligomer **9** are linked by a single carbon-carbon bond, resulting in fewer degrees of rotational freedom, which may result in a reduced entropic penalty for minor-groove binding. Undoubtedly, much of the favorable energetics for complexation with DNA for all these molecules is a result of differential solvation. The oligomer's large hydrophobic surface may result in increased favorable van der Waals interactions with the walls of the minor groove.

### C.5 Conclusion

Hairpin polyamides containing the No-Hz and Ct-Hz dimer caps at the polyamide N-terminus are able to target 5'-GT-3' and 5'-TT-3' sequences with good affinity and specificity and represent new recognition elements for the minor groove of DNA. The No-Hz and Ct-Hz dimer caps represent attempts to broaden heterocycle designs beyond single base pair interactions. In addition, the development of the No-Hz cap has allowed for the design of a DNA binding molecule, which in a formal sense is no longer a polyamide, hence the term programmable oligomer. We are encouraged by the fact that this oligomer demonstrates excellent affinity for DNA while exhibiting good levels of specificity. We hope to apply these new heterocycles to the targeting of biologically relevant sequences in the context of integrating artificial transcription factors with living biological systems.

### C.6 Notes and References

1. Darnell, J. E. Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* **2002**, *2*, 740-748. (b) Pandolfi, P. P. Transcription therapy for cancer. *Oncogene* **2001**, *20*, 3116-3127. (c) Dervan, P. B.; Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.
2. (a) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. Netropsin, A new antibiotic produced by a streptomyces. *J. Am. Chem. Soc.* **1951**, *73*, 341-343. (b) Arcamone, F. N. V.; Penco,

- S.; Orezzi, P.; Nicoletta, V.; Pirelli, A. Structure and synthesis of Distamycin A. *Nature* **1964**, *203*, 1064-1065.
3. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376-1380. (b) Pelton, J. G.; Wemmer, D. E. Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723-5727. (c) Wemmer, D. E. Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 439-461. (d) Buchmueller, K. L.; Staples, A. M.; Howard, C. M.; Horick, S. M.; Uthe, P. B.; Minh Le, N.; Cox, K. K.; Nguyen, B.; Pacheco, K. A. O.; Wilson, D. W.; Lee, M. Extending the language of DNA molecular recognition by polyamides: unexpected influence of imidazole and pyrrole arrangement on binding affinity and specificity. *J. Am. Chem. Soc.* **2004**, *127*, 742-750. (e) Baraldi, P. G.; Bovero, A.; Fruttarolo, F.; Preti, D.; Tabrizi, M. A.; Pavani, M. G.; Romagnoli, R. DNA minor groove binders as potential antitumor and antimicrobial agents. *Med. Res. Rev.* **2004**, *24*, 475-528. (f) Reddy, P. M.; Toporowski, J. W.; Kahane, A. L.; Bruice, T. C. Recognition of a 10 base pair sequence of DNA and stereochemical control of the binding affinity of chiral hairpin polyamide-Hoechst 33258 conjugates. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5531-5536.
4. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **1998**, *391*, 468-471.
5. (a) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science* **1998**, *282*, 111-115. (b) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. Structural effects of DNA sequence on TA recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **2000**, *295*, 557-567.
6. (a) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Sequence selectivity of 3-hydroxypyrrole/pyrrole ring pairings in the DNA minor groove. *J. Am. Chem. Soc.* **1999**, *121*, 11621-11629. (b) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. Affinity and specificity of multiple hydroxypyrrole/pyrrole ring pairings for coded recognition of DNA. *J. Am. Chem. Soc.* **1999**, *121*, 260-261.
7. (a) Hays, F. A.; Teegarden, A.; Jones, Z. J. R.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Shing Ho, P. How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7157-7162. (b) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham 3rd, T. E.; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.* **2004**, *87*, 3799-3813. (c) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham 3rd, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.* **2005**, *89*, 3721-3740. (d) Wu, H.; Crothers, D. M. The locus of sequence-directed and protein-induced DNA bending. *Nature* **1984**, *308*, 509. (e) Steitz, T. A. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* **1990**, *23*, 205. (f) Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2930. (g) Paoletta, D. N.; Palmer, R.; Schepartz, A. DNA targets for certain bZIP proteins distinguished by an intrinsic

bend. *Science* **1994**, *264*, 1130. (h) Kahn, J. D.; Yun, E.; Crothers, D. M. Detection of localized DNA flexibility. *Nature* **1994**, *368* 163. (h) Geierstanger, B. H.; Wemmer, D. E. Complexes of the minor groove of DNA. *Annu. Rev. Biophys. Biochem. Struct.* **1995**, *24*, 463. (i) Hansen, M. R.; Hurley, L. H. Pluramycins. Old Drugs Having Modern Friends in Structural Biology. *Acc. Chem. Res.* **1996**, *29*, 249. (j) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226. (k) Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B. Toward an understanding of the chemical etiology for DNA minor-groove recognition by polyamides. *Helv. Chim. Acta* **2002**, *85*, 4485-4517.

8. Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B. Shape selective recognition of T•A base pairs by hairpin polyamides containing N-terminal 3-methoxy (and 3-chloro) thiophene residues. *Bioorg. Med. Chem.* **2003**, *11*, 4333-4340.

9. Briehn, C. A.; Weyermann, P.; Dervan, P. B. Alternative heterocycles for DNA recognition: the benzimidazole/imidazole pair. *Chem. Eur. J.* **2003**, *9*, 2110-2122.

10. (a) Renneberg, D.; Dervan, P. B. Imidazopyridine/Pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc.* **2003**, *125*, 5707-5716. (b) Marques, M. A.; Doss, R. M.; Foister, S.; Dervan, P. B. Expanding the repertoire of heterocycle ring pairs for programmable minor groove DNA recognition. *J. Am. Chem. Soc.* **2004**, *126*, 10339-10349.

11. For down regulation of an endogenous gene in cell culture by polyamides see: Olenyuk, B. Z.; Zhang, G.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G. Jr.; Dervan, P. B. Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16768-16773.

12. Heterocycle pair refers to cofacial stack (noncovalent interaction) whereas dimer refers to two covalently attached heterocycles.

13. Baird, E. E.; Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

14. Urbach, A. R.; Dervan, P. B. Toward rules for 1: 1 polyamide: DNA recognition. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4343-4348.

15. Trauger, J. W.; Dervan, P. B. Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. *Methods Enzymol.* **2001**, *340*, 450-466.

16. Im-Hp-Py-Py- $\gamma$ -Im-Hp-Py-Py- $\beta$ -Dp targets 5'-WGTACW-3' with  $K_a = 7.0 \times 10^8 \text{ M}^{-1}$  and Im-Hz-Py-Py- $\gamma$ -Im-Hz-Py-Py- $\beta$ -Dp targets 5'-WGTACW-3' with  $K_a = 4.6 \times 10^8 \text{ M}^{-1}$  where W = A or T<sup>10b</sup>

17. Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B. Structure of a beta-alanine-linked polyamide bound to a full helical turn of purine tract DNA in the 1: 1 motif. *J. Mol. Biol.* **2002**, *320*, 55-71.

18. Suzuki, M.; Iwasaki, T.; Miyoshi, M.; Okumura, K.; Matsumoto, K. Synthesis of amino acids and related compounds. 6. New convenient synthesis of alpha-C-acylamino acids and alpha-amino ketones. *J. Org. Chem.* **1973**, *38*, 3571-3575.

## C.7 Supplemental Information

### C.7.1 General

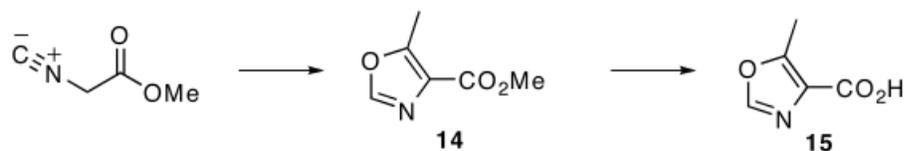
*N,N*-dimethylformamide (DMF), *N,N*-diisopropylethylamine (DIEA), thiophenol (PhSH), 3-dimethylamino-1-propylamine (Dp), Triethylamine (TEA), and thiourea were purchased from Aldrich. Boc- $\beta$ -alanine-(4-carboxylaminomethyl)benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- $\beta$ -Pam-resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 4-dimethylaminopyridine (DMAP), and Boc- $\beta$ -alanine were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All other solvents were reagent grade from EMD Chemicals. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Precoated silica gel plates 60 F<sub>254</sub> for TLC and silica gel 60 (40  $\mu$ m) for flash chromatography were from Merck. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [ $\gamma$ -<sup>32</sup>P] triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) and DNaseI (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA polymerase was from Perkin-Elmer and used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO, and bromophenol blue was from Acros. All reagents were used without further purification. NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d*<sub>6</sub> or CDCl<sub>3</sub>, with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett-Packard model 8452A diode-array spectrophotometer. High-resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the California Institute of Technology. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was conducted at the Mass Spectroscopy Laboratory at the California Institute of Technology.

### C.7.2 Heterocycle Synthesis

The synthesis of compounds **4** (Boc-Py-OBt), **5** (Boc-Im-OH), **10** (Boc-Py-Bi-OH), **11** (Boc-Im-HzOMe-OH), **19** (Ct-OH), **10**, **22** (Im-HzOMe-OH), and **16** (aryl diamine) have previously been reported.<sup>8,10a</sup>

#### C.7.2.1 Methyl 5-methyl-1,3-oxazole-4-carboxylate (No-OMe) (**14**)

Compound **14** was prepared following an exact preparation reported by Suzuki et al.<sup>18</sup> and purified



**Scheme C.2** Synthesis of **15**.

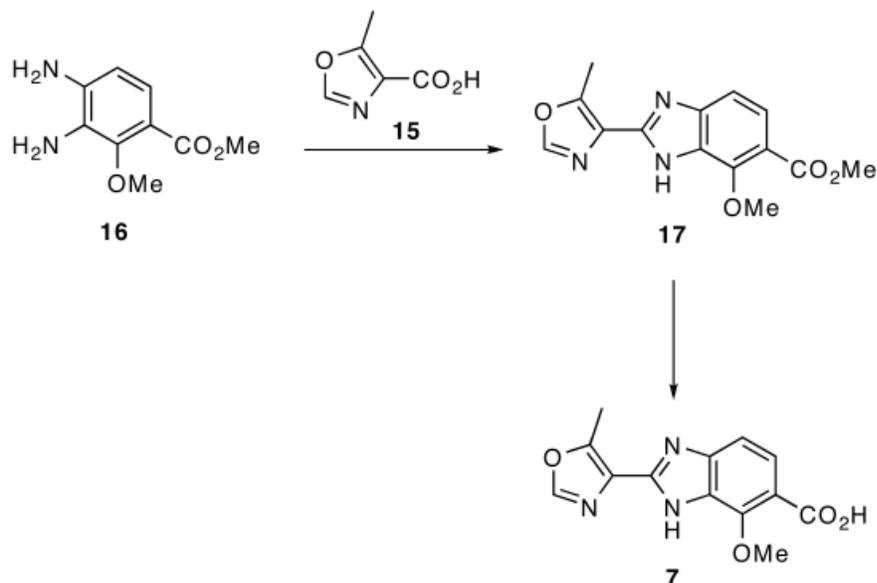
by normal phase column chromatography. According to the Suzuki procedure a mixture of methyl- $\alpha$ -isocyanoacetate (3.0 g, 30.2 mmol) and DBU (4.5 g, 30.2 mmol) in dry THF (40 mL) at 10°C, was treated with acetic anhydride (3.06 g, 30 mmol) in dry THF (10 mL) dropwise. The reaction was allowed to warm to room temperature and stirred for 10 h. The solvent was removed by rotoevaporation and water (100 mL) was added. The mixture was extracted with EtOAc (2 x 100 mL). The organic layers were combined, dried over sodium sulfate, filtered, and concentrated in vacuo to provide **14** as a crude amber oil. The oil was subjected to column chromatography using (3:2 Hex/EtOAc) to provide **14** (2.73 g, 64% Yield) as a crystalline white solid. TLC (3:2 Hex/EtOAc)  $R_f$  0.5;  $^1\text{H NMR}$  (DMSO- $d_6$ ) 8.32 (s, 1H), 3.77 (s, 3H), 2.55 (s, 3H);  $^{13}\text{C}$  (DMSO- $d_6$ ) 161.9, 156.1, 150.3, 126.2, 51.5, 11.5; HR-MS (EI): calculated for  $\text{C}_6\text{H}_7\text{NO}_3$ : 141.0426; found: 141.0427.

*C.7.2.2 5-Methyl-1,3-oxazole-4-carboxylic acid (No-OH) (15)*

A mixture of **14** (1 g, 7.08 mmol), 1N NaOH (10 mL) and MeOH (5 mL) was stirred at 40°C for 4 h. The MeOH was removed by rotoevaporation and the pH adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **15** (738 mg, 82% Yield) as a fibrous white solid. TLC (3:2 EtOAc/Hex +10% AcOH)  $R_f$  0.4;  $^1\text{H NMR}$  (DMSO- $d_6$ ) 8.27 (s, 1H), 2.45 (s, 3H);  $^{13}\text{C}$  (DMSO- $d_6$ ) 163.0, 155.6, 150.1, 145.6, 11.6; HR-MS (EI): calculated for  $\text{C}_5\text{H}_5\text{NO}_3$ : 127.0269; found: 127.0268.

*C.7.2.3 Methyl 7-methoxy-2-(5-methyl(1,3-oxazole-4-yl))benzimidazole-6-carboxylate (NoHz(OMe)OMe) (17)*

To a solution of **15** (0.3 g, 2.36 mmol) in DMF (4 mL) was added DIEA (915 mg, 1.23 mL, 7.08 mmol) and HBTU (895 mg, 2.36 mmol). The mixture was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **16** (463 mg, 2.36 mmol). The reaction was then heated to 35°C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a separatory funnel containing water (200 mL). The water was then extracted with



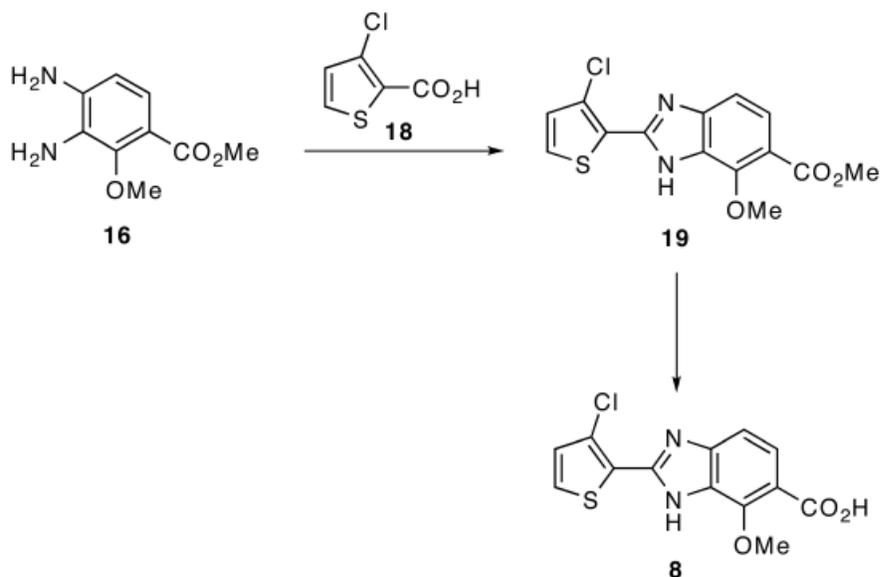
### Scheme C.3 Synthesis of **7**.

EtOAc (2 x 100 mL). The organic layer was dried over anhydrous sodium sulfate and concentrated in vacuo to provide a crude solid, which was dissolved in glacial acetic acid (5 mL) and heated to 90°C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (4:1 EtOAc/Hex) to provide **17** as a thin film. Addition of hexanes to the film, followed by rotoevaporation and drying under high vacuum provided **17** as a white solid (379 mg, 56% Yield). TLC (4:1 EtOAc/Hex)  $R_f$  0.75;  $^1\text{H}$  NMR (DMSO- $d_6$ ) 8.49 (s, 1H), 7.51 (d, 1H,  $J = 8.7$  Hz), 7.16 (d, 1H,  $J = 8.7$  Hz), 4.34 (s, 3H), 3.77 (s, 3H) 2.78 (s, 3H);  $^{13}\text{C}$  (DMSO- $d_6$ ) 166.7, 151.0, 149.3, 145.5, 139.0, 135.3, 125.6, 124.8, 114.2, 105.3, 60.8, 51.6, 11.3; HR-MS (EI): calculated for  $\text{C}_{14}\text{H}_{13}\text{N}_3\text{O}_4$ : 287.0906; found: 287.0914.

#### C.7.2.4 7-methoxy-2-(5-methyl(1,3-oxazole-4-yl))benzimidazole-6-carboxylic acid (NoHz(OMe) OH) (**7**)

A mixture of **17** (200 mg, 0.69 mmol), DMSO (0.8 mL), and 1N NaOH (4 mL) was stirred at 23 °C for 3 h. The solution was diluted with 8 mL of  $\text{H}_2\text{O}$  and the pH was adjusted to pH = 2 with 1N HCl (4.4 mL). The precipitate was filtered and washed with diethyl ether. The crude material was purified by silica gel chromatography (1:1 EtOAc/Hexanes + 2% AcOH to 3:1 EtOAc/Hexanes + 2% AcOH) to provide **7** (156 mg, 82% Yield) as a white solid after drying under high vacuum. TLC (3:1 EtOAc/Hex +2% AcOH)  $R_f$  0.55;  $^1\text{H}$  NMR (DMSO- $d_6$ ) 13.12 (s, 1H), 12.37 (s, 1H), 8.51 (s, 1H), 7.52 (d, 1H,  $J = 8.5$  Hz), 7.15 (d, 1H,  $J = 8.5$  Hz), 4.34 (s, 3H), 2.78 (s, 3H);  $^{13}\text{C}$  (DMSO- $d_6$ ) 167.8, 151.1, 150.9, 149.3, 145.5, 139.0, 135.5, 125.8, 125.2, 115.2, 105.4, 60.9, 11.4; HR-MS

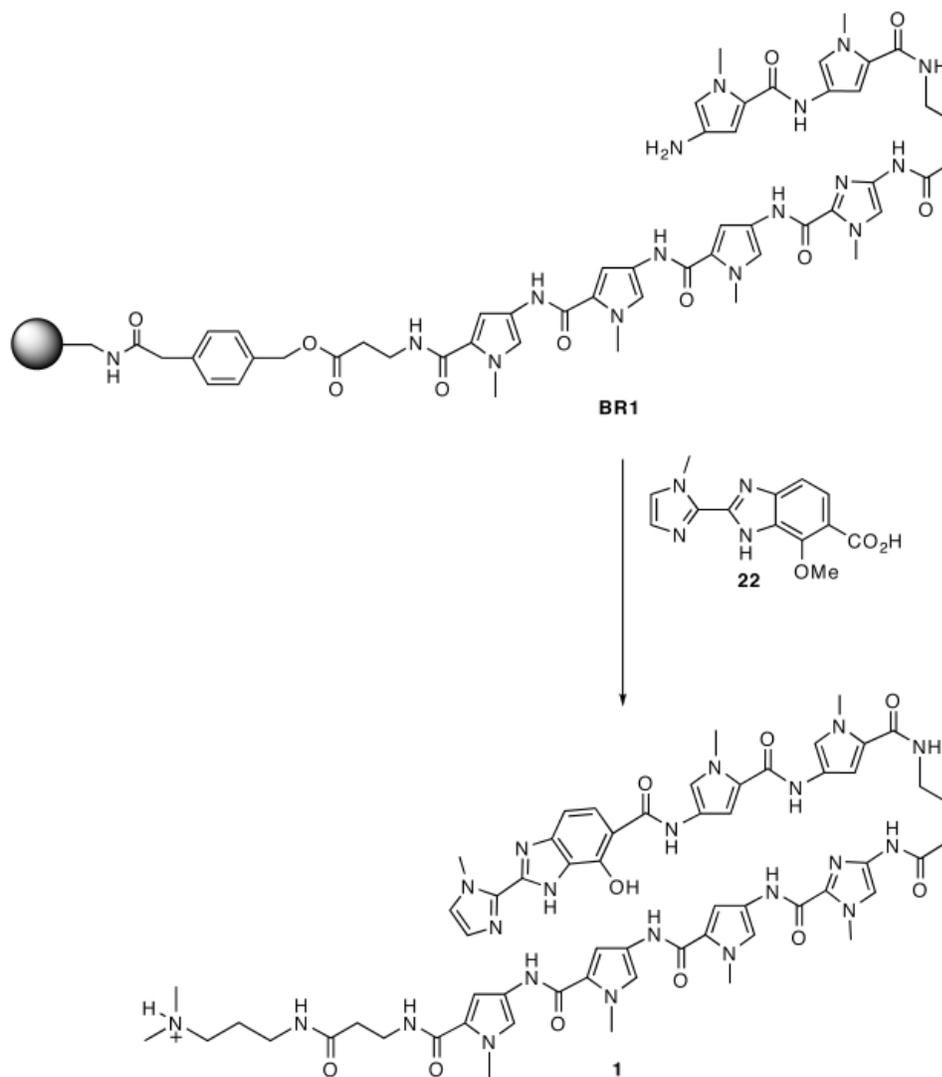
(FAB): calculated for  $C_{13}H_{12}N_3O_4$ : 274.0828; found: 274.0833  $[M+H]^+$ .



**Scheme C.4** Synthesis of **8**.

#### C.7.2.5 *Ct-Hz(OMe)OMe (19)*

To a solution of **18** (294 mg, 1.80 mmol) in DMF (5 mL) was added DIEA (0.348 mL, 1.98 mmol) and HBTU (650 mg, 1.72 mmol). The mixture was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **16** (300 mg, 1.80 mmol). The reaction was then heated to 35°C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a Falcon tube containing cold water (40 mL) resulting in a cloudy precipitate. The Falcon tube was centrifuged at 14000 rpm for 10 min, the mother liquor decanted, and the precipitate dried under high vacuum. After drying, the crude solid was dissolved in glacial acetic acid (5 mL) and heated to 90 °C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (3:2 EtOAc/Hex) to provide **19** as an off-white solid (647 mg, 57% Yield). TLC (3:2 EtOAc/Hex)  $R_f$  0.5;  $^1H$  NMR (DMSO- $d_6$ ) 12.85 (s, 1H), 7.89 (d, 1H,  $J = 5.3$  Hz), 7.56 (d, 1H,  $J = 8.5$  Hz), 7.27 (d, 1H,  $J = 8.2$ ), 7.26 (d, 1H,  $J = 5.3$ ), 4.30 (s, 3H), 3.80 (s, 3H);  $^{13}C$  (DMSO- $d_6$ ) 166.6, 151.2, 144.0, 139.8, 134.8, 129.7, 129.4, 126.3, 125.6, 123.2, 114.8, 106.1, 61.2, 51.8; HR-MS (EI): calculated for  $C_{14}H_{11}N_2O_3S$ : 322.0179; found: 322.0171.

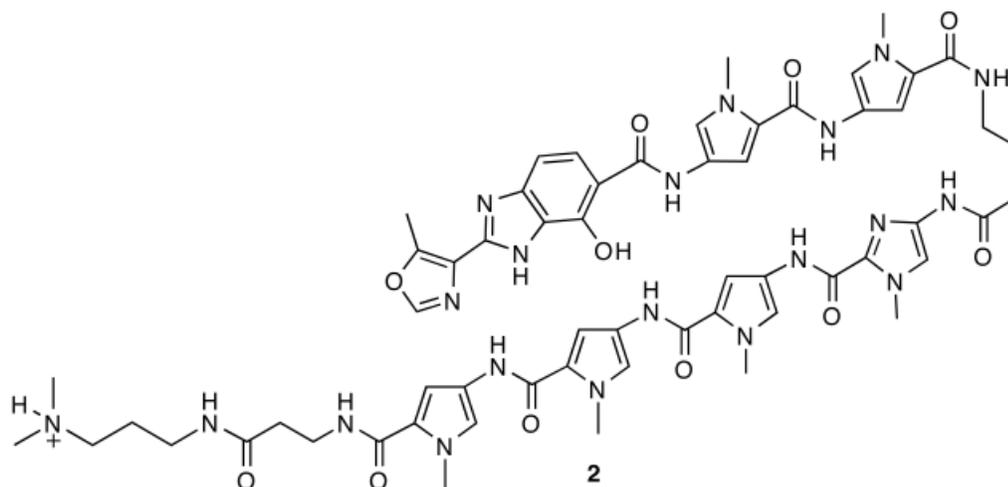


**Scheme C.5** Synthesis of **1**.

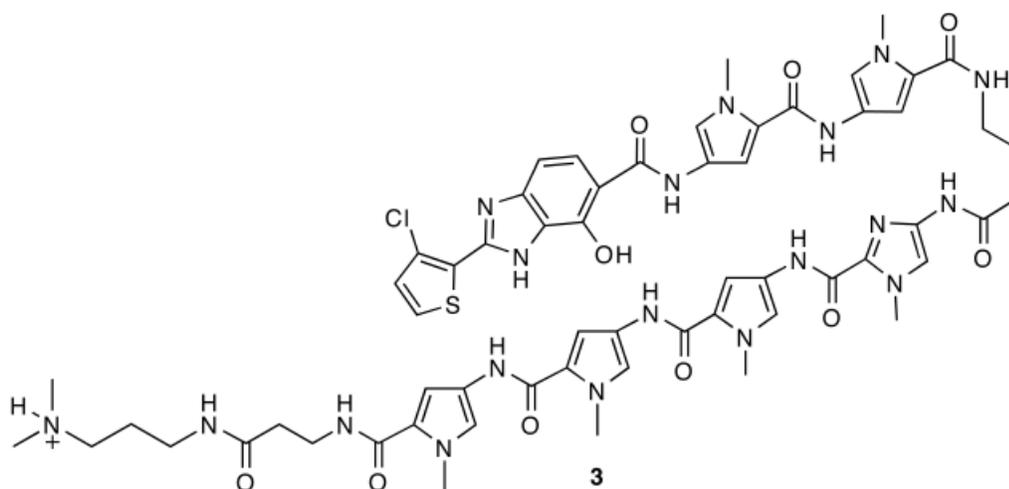
#### C.7.2.6 Ct-Hz(OMe)OH (**8**)

A mixture of **19** ( 1.8 g, 5.58 mmol), 1N NaOH (36 mL) and MeOH (27 mL) was stirred at 45°C for 24 h. The MeOH was removed by rotoevaporation and the pH adjusted to pH = 5.5 with 1N HCl, resulting in a precipitate. The mixture was spun down in a centrifuge (5 min at 14,000 rpm). The tube was decanted leaving a white solid that was dried under high vacuum to provide **8** (2.6 g, Quant. Yield) as a light yellow solid. TLC (3:2 EtOAc/Hex + 10% AcOH)  $R_f$  0.5;  $^1\text{H}$  NMR (DMSO- $d_6$ ) 13.00 (s, 1H), 7.89 (d, 1H,  $J = 5.3$  Hz), 7.57 (d, 1H,  $J = 8.5$  Hz), 7.30 (d, 1H,  $J = 8.5$ ), 7.25 (d, 1H,  $J = 5.3$  Hz), 4.28 (s, 3H);  $^{13}\text{C}$  (DMSO- $d_6$ ) 167.7, 151.1, 143.9, 139.7, 134.9, 129.6, 129.3, 126.4, 125.7, 123.2, 115.9, 106.1, 61.2; HR-MS (EI): calculated for  $\text{C}_{13}\text{H}_9\text{N}_2\text{O}_3\text{Cl}$ : 308.0022; found: 308.0033.

## C.7.3 Polyamide Synthesis



Scheme C.6 Compound 2.



Scheme C.7 Compound 3.

Polyamides were synthesized from Boc- $\beta$ -alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid phase protocols.<sup>4</sup>The synthesis of batch resin BR1 (H<sub>2</sub>N-Py-Py- $\gamma$ -Im-Py-Py-Py- $\beta$ -Pam) has previously been reported.<sup>8</sup>

C.7.3.1 Im-Hz-Py-Py- $\gamma$ -Im-Py-Py-Py- $\beta$ -Dp (1)

**22** (Im-HzOMe-OH) (25 mg, 88.5  $\mu$ mol) was incorporated by activation with HBTU (32 mg, 84  $\mu$ mol), DIEA (23 mg, 31  $\mu$ l, 177  $\mu$ mol) and DMF (250  $\mu$ l). The mixture was allowed to stand for

15 min at room temperature and then added to the reaction vessel containing base resin BR1 H<sub>2</sub>N-Py-Py-γ-Im-Py-Py-Py-β-Pam. Coupling was allowed to proceed for 12 h at room temperature. The resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide Im-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**1**) (0.9 mg, 2.4 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS: calculated for C<sub>59</sub>H<sub>70</sub>N<sub>21</sub>O<sub>10</sub>: 1232.56; found 1232.55 [M+H]<sup>+</sup>.

#### C.7.3.2 No-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**2**)

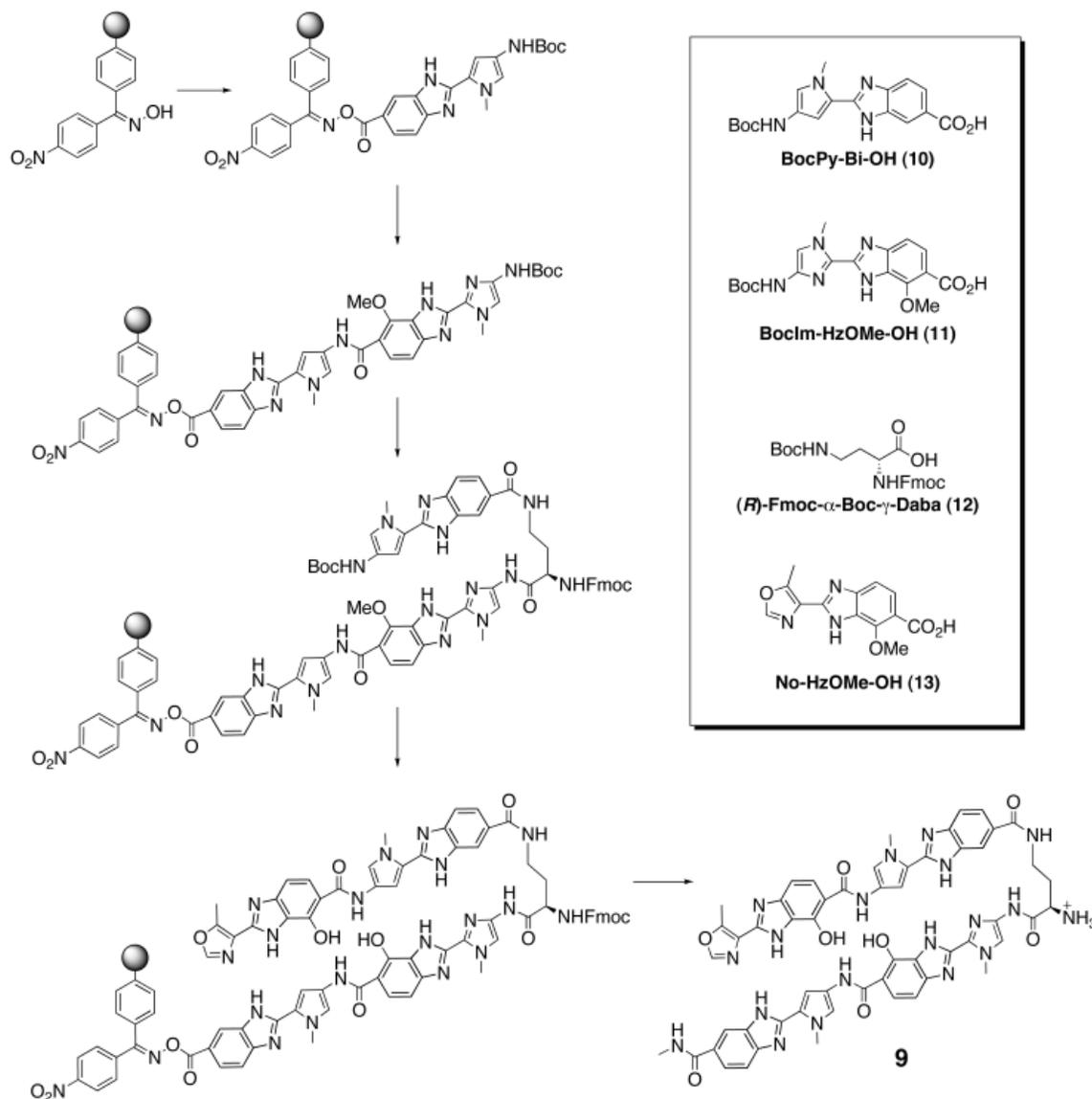
**7** (No-HzOMe-OH) was incorporated as described above for Im-Hz-OH (polyamide **1**) to provide No-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**2**) (1.5 mg, 4.1 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS: calculated for C<sub>59</sub>H<sub>69</sub>N<sub>20</sub>O<sub>11</sub>: 1233.55; found 1233.54 [M+H]<sup>+</sup>.

#### C.7.3.3 Ct-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**3**)

**8** (Ct-HzOMe-OH) was incorporated as described above for Im-Hz-OH (polyamide **1**) to provide Ct-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**3**) (1.1 mg, 2.9 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS: calculated for C<sub>59</sub>H<sub>67</sub>ClN<sub>19</sub>O<sub>10</sub>S: 1269.47; found 1269.47 [M+H]<sup>+</sup>.

#### C.7.4 Deprotection of the *O*-Methyl-Protected Polyamides

*O*-Methyl protected polyamides were cleaved from resin, purified, deprotected and subject to further purification using the following general procedure. Upon completion of solid phase synthesis, 3-dimethylamino-1-propylamine (Dp) (500 μL) was added to the synthesis vessel containing the resin (50 mg). The mixture was allowed to stand for 2 h at 85°C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC and lyophilized to provide polyamides containing the *O*-methyl protected hydroxybenzimidazole unit (-HzOMe-) as a dry solid. The polyamides were then dissolved in DMF (200 μl) and added to a suspension of sodium hydride (40 mg, 60% oil dispersion) and thiophenol (200 μl) in DMF (400 μl) that was pre-heated for 5 min at 85°C. The mixture was heated for 2 h at 85°C. The mixture was then cooled to 0 °C and 20% TFA (7.0 mL) was added. The aqueous layer was washed three times with diethyl ether (8 mL) and then diluted to a total volume of 9.5 mL using 0.1% TFA. The mixture was then purified by reverse-phase HPLC to give the deprotected



**Scheme C.8** Synthesis of **9-13**.

Hz-containing polyamides **1**, **2**, and **3**.

#### C.7.5 Oligomer **9** Synthesis

Oligomer **9** was prepared using Kaiser oxime resin (0.65 mequiv/gram) from Nova Biochem. Oxime resin (0.1 grams), was added to a solid phase synthesis vessel. The resin was then washed with DCM (15 mL) followed by DMF (15 mL). In a separate vessel was added Boc-PyBi-OH (**10**) (60 mg, 0.168 mmol), HBTU (61 mg, 0.159 mmol), DIEA (43 mg, 60  $\mu$ L, 0.336 mmol), and DMF (400  $\mu$ L). The mixture was vortexed and allowed to activate for 20 minutes at room temperature. The

activated mixture was then added to the resin. The reaction vessel was shaken at room temperature overnight. The reaction vessel was then drained and the resin washed with DMF (15 mL), followed by DCM (15 mL). The resin was then washed with a solution of 25% TFA in DCM (20 mL), followed by shaking with 25% TFA in DCM at room temperature for 25 min to deprotect the Boc-protected alkyl amine, followed by shaking at room temperature for 25 min to deprotect the Boc-protected amine of pyrrole. In a separate vessel, a mixture of Boc-ImHz-OH (**11**) was activated using HBTU, DIEA, and DMF as described above. The reaction vessel containing the resin was then drained and washed with DCM (15 mL) and 10% DIEA in DMF. The activated Boc-ImHz-OBt dimer was then added to the vessel containing (Resin-BiPy-NH<sub>2</sub>). The mixture was shaken at room temperature for 2.5 h. The reaction vessel was then drained and washed with DMF (15 mL) and DCM (15 mL). A solution of 50% TFA in DCM was then washed over the resin (20 mL), followed by shaking at room temperature for 25 minutes. In a separate vessel, a mixture of  $\alpha$ -Fmoc- $\gamma$ -Boc-(*R*)-2,4-diaminobutyric acid, ((*R*)-Fmoc- $\alpha$ -Boc- $\gamma$ -Daba, and (**12**) was activated using HBTU, DIEA, and DMF as described above. The reaction vessel containing the resin was then drained and washed with DCM (15 mL) and 10% DIEA in DMF. The activated (*R*)-Fmoc- $\alpha$ -Boc- $\gamma$ -Daba-OBt monomer was then added to the vessel containing (Resin-BiPy-HzIm-NH<sub>2</sub>). The mixture was shaken for 2 h at room temperature. The reaction vessel was then drained and the resin washed with DMF (15 mL), followed by DCM (15 mL). The resin was then washed with a solution of 25% TFA in DCM (20 mL), followed by shaking with 25% TFA in DCM at room temperature for 25 min to deprotect the Boc-protected alkyl amine. In a separate vessel, a mixture of Boc-PyBi-OH (**10**) was activated using HBTU, DIEA and DMF as described above. The reaction vessel containing the resin was then drained and washed with DCM (15 mL) and 10% DIEA in DMF. The activated Boc-PyBi-OBt dimer was then added to the vessel containing (Resin-BiPy-HzIm- $\gamma$ Daba-NH<sub>2</sub>). The mixture was shaken at room temperature for 2.5 h. The reaction vessel was then drained and the resin washed with DMF (15 mL), followed by DCM (15 mL). The resin was then washed with a solution of 25% TFA in DCM (20 mL), followed by shaking with 25% TFA in DCM at room temperature for 25 min to deprotect the Boc-protected alkyl amine, followed by shaking at room temperature with 25% TFA in DCM for 25 min to deprotect the Boc-protected amine of pyrrole. In a separate vessel, a mixture of NoHzOMe-OH (**13**) was activated using HBTU, DIEA and DMF as described above for (**10**). The reaction vessel containing the resin was then drained and washed with DCM (15 mL) and 10% DIEA in DMF. The activated NoHzOMe-OBt dimer was then added to the vessel containing (Resin-BiPy-HzIm- $\gamma$ Daba-BiPyNH<sub>2</sub>). The mixture was shaken at room temperature for

2.5 h. The reaction vessel was then drained and washed with DMF (15 mL) and DCM (15 mL) to provide (Resin-BiPy-HzIm- $\gamma$ Daba-BiPy-HzNo).

#### *C.7.6 O-Methyl Deprotection*

To the synthesis vessel containing (Resin-BiPy-HzIm- $\gamma$ Daba-BiPy-HzNo) was added DCM (800  $\mu$ L) and BCl<sub>3</sub> (400  $\mu$ L), (2M in heptanes). The mixture was shaken at room temperature for 2.5 h. The vessel was then drained and washed with DCM (15 mL). Deprotection was determined to be complete by analytical HPLC

#### *C.7.7 Cleavage From Resin*

Following O-Methyl deprotection, DCM (200  $\mu$ L) and MeNH<sub>2</sub> (1 mL) (2M in THF) was added to the synthesis vessel. The mixture was then shaken at 35°C for 12 h. The filtrate was collected from the synthesis vessel and the organics removed in vacuo. The remaining residue was dissolved in 0.1% TFA and purified using preparatory reverse phase HPLC. Lyophilization of the appropriate fractions provided oligomer **9** as a fine powder solid. MALDI-TOF-MS: calcd for C<sub>55</sub>H<sub>49</sub>N<sub>19</sub>O<sub>8</sub>: 1103.40; found 1103.41 [M+H]<sup>+</sup>.

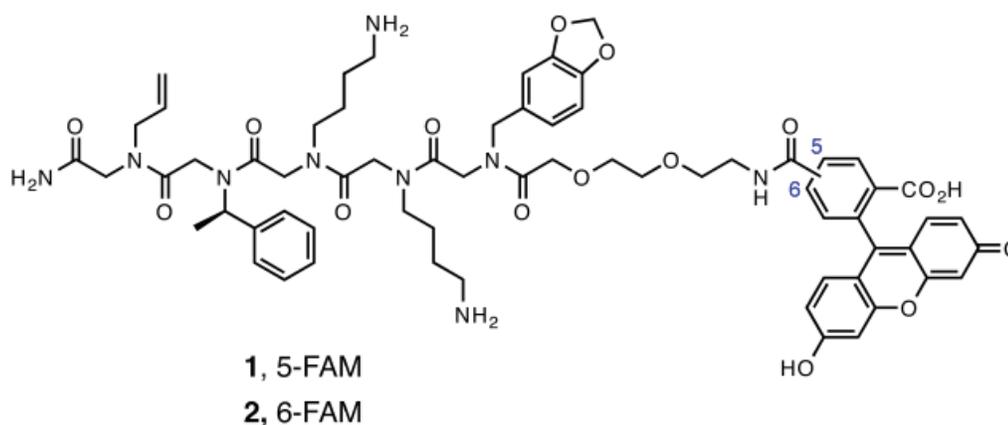
## **Appendix D: Peptoid Cell Uptake Studies**

**Abstract**

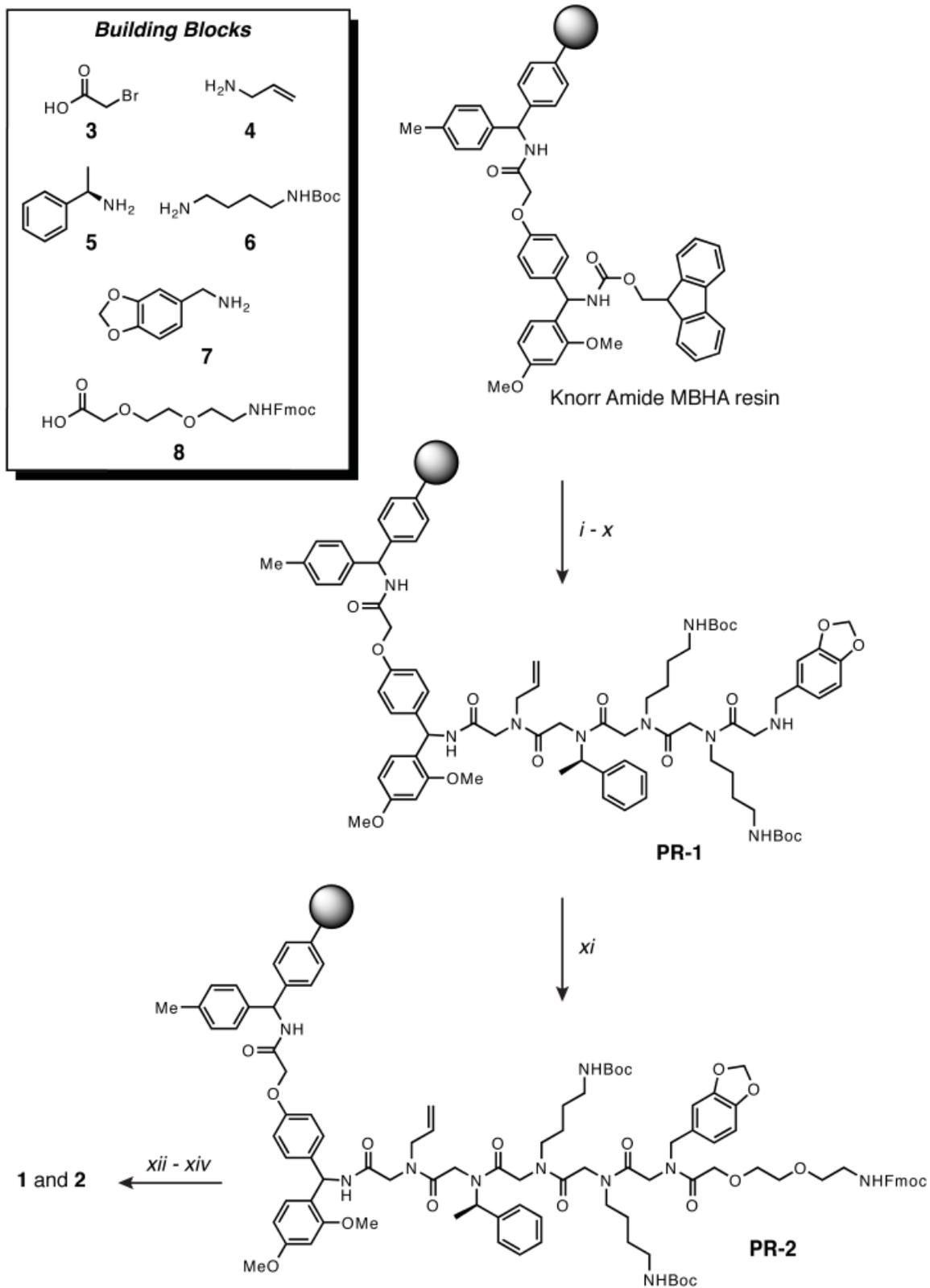
Peptoids 5-FAM (**1**) and 6-FAM (**2**), previously reported by the Kodadek group, were synthesized and their cell uptake properties evaluated in HeLa cells using confocal laser scanning microscopy. A dual imaging laser system was used to image the peptoids and a DNA-binding nuclear stain simultaneously. The results demonstrate that the peptoids 5-FAM (**1**) and 6-FAM (**2**) are not cell permeable in HeLa cells.

## D.1 Summary

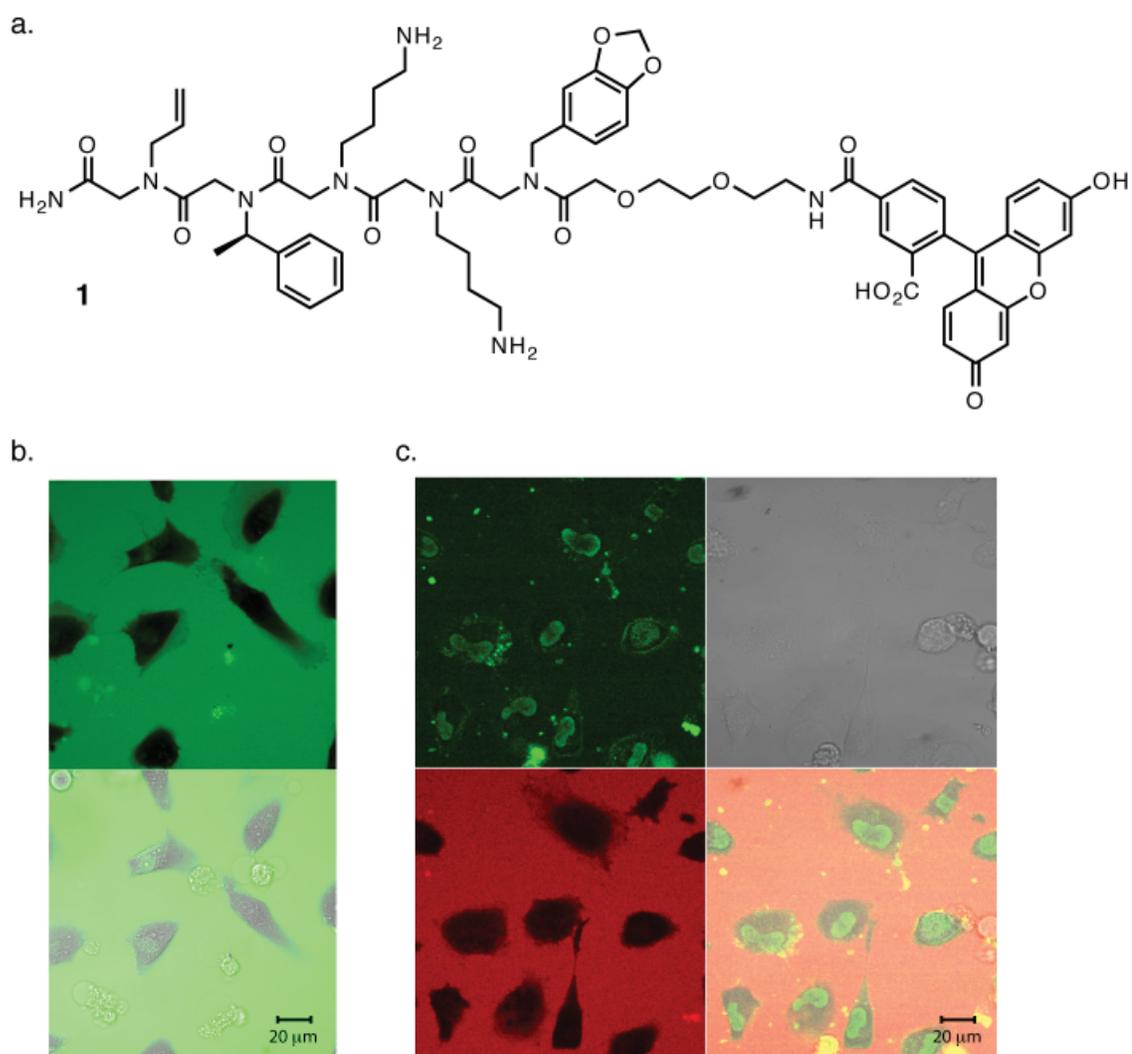
Peptoids are a class of *N*-alkylated poly-glycine oligomers that have recently been exploited for the rapid synthesis of protein-targeted combinatorial libraries in an effort to discover new bio-active compounds with superior cell permeability and protease sensitivity as compared to traditional peptide scaffolds.<sup>1-3</sup> Recently, the Kodadek group has demonstrated that these compounds are often cell permeable and can be utilized as molecular recognition domains.<sup>4</sup> A peptoid was shown to be effective as a transcriptional activation domain for GST-KIX in a HeLa cell reporter gene assay.<sup>5</sup> The peptoid domain which bound a GST-KIX fusion protein with a  $K_d$  of 11.6  $\mu\text{M}$  was discovered from a library of 50,000 members screened against the murine CREB core KIX domain. A carboxyfluoresceinated version of this peptoid (mixture of **1** and **2**, Figure D.1) was used for KIX binding studies in comparison to a scrambled peptoid sequence, which was shown not to bind the GST-KIX domain. In an effort to evaluate the potential use of this peptoid activation domain for use with different polyamide DNA binding scaffolds, the exact peptoid-fluorophore conjugates (**1** and **2**, Figure D.1) from the Kodadek study were synthesized independently (not as a mixture) as shown in Figure D.2. The cell uptake properties of these peptoid-FAM conjugates were assessed using confocal laser scanning microscopy and dual imaging studies with DNA binding dyes were performed to provide unambiguous location of the cell nuclei. These studies were performed in HeLa cells and show that the peptoid was completely excluded from the cell interior as shown in Figures D.3 and D.4.



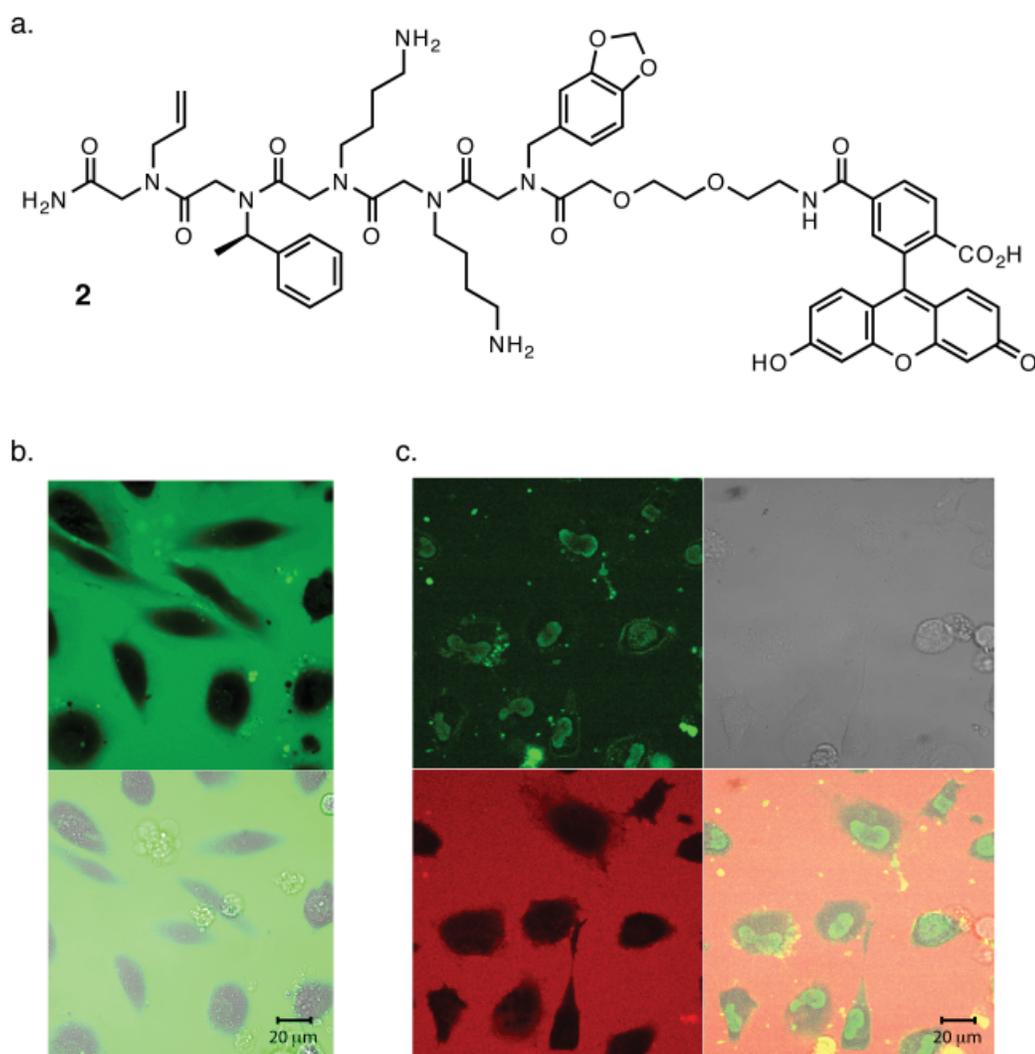
**Figure D.1** Compound **1**, 5-FAM, and **2**, 6-FAM.



**Figure D.2** Synthesis of compound **1** and **2**. For details, see Section D.2 Experimental.



**Figure D.3** HeLa cell uptake studies for compound **1** (5-FAM) using  $2\mu\text{M}$  concentration. Dual image with Hoechst 33342. (2-Photon Laser 1 = 810 nm, 5% Power, BP480 - 520 nm filter.).



**Figure D.4** HeLa cell uptake studies for compound **2** (6-FAM) using 2 $\mu$ M concentration. Dual image with Hoechst 33342. (2-Photon Laser 1 = 810 nm, 5% Power, BP480 - 520 nm filter.)

## D.2 Experimental

### D.2.1 Materials

#### D.2.1.1 Resin

Knorr Amide MBHA resin, Nova Biochem, 01-64-0459, A33927

Resin loading = 0.78 mmol/g

Amount = 200 mg

#### D.2.1.2 Stock solution preparation

Stock solution A

2 mL of piperidine in 8 mL of DMF.

Stock solution B

3.7 g of bromoacetic acid (**3**) in 14 mL of DMF.

Stock solution C

7 mL of *N,N'*-diisopropylcarbodiimide in 7 mL of DMF.

#### *D.2.1.3 Amine solution preparation*

Solution M1

300  $\mu$ L of allylamine (**4**) in 1.7 mL DMF.

Solution M2

510  $\mu$ L of (*R*)-(+)- $\alpha$ -methylbenzylamine (**5**) in 1.49 mL DMF.

Solution M3

1.5 g (1.53 mL) of *N*-Boc-1,4-butane diamine (**6**) in 2.47 mL DMF.

Solution M4

1.5 g (1.53 mL) of *N*-Boc-1,4-butane diamine (**6**) in 2.47 mL DMF.

Solution M5

500  $\mu$ L of piperonylamine (**7**) in 1.5 mL DMF.

#### *D.2.2 General bromoacetic acid addition procedure*

Resin was treated with 2 mL of stock solution B followed by 2 mL of stock solution C. Next, the synthesis vessel was capped and microwaved for 15 sec on the low power setting followed by shaking and another 15 sec in the microwave. Next, the resin synthesis vessel was drained and washed with DMF (4 x 5 mL) followed by a final wash with 10 mL of anhydrous DMF.

#### *D.2.3 General amine addition procedure*

Resin was treated with the appropriate amine solution and agitated for 15 sec followed by 15 sec in the microwave on the low power setting. The reaction mixture was agitated for another 15 sec and microwaved again for 15 sec on the lower power setting followed by draining of the vessel and washed with DMF (4 x 5 mL). Finally, a wash with 10 mL of anhydrous DMF was performed.

#### *D.2.4 Peptoid synthesis procedure (Synthesis of Resin PR-1)*

A solid phase synthesis vessel was charged with 200 mg of Knorr Amide MBHA resin and swelled

in DMF for 20 min with agitation. The resin was then washed 4 times with DMF, treated with 20% piperidine in DMF for 20 min (2 mL x 20 min each), washed with DMF (8 x 5 mL), and washed with 10 mL of anhydrous DMF.

1. General bromoacetic acid addition procedure.
2. General amine addition procedure using solution M1.
3. General bromoacetic acid addition procedure.
4. General amine addition procedure using solution M2.
5. General bromoacetic acid addition procedure.
6. General amine addition procedure using solution M3.
7. General bromoacetic acid addition procedure.
8. General amine addition procedure using solution M4.
9. General bromoacetic acid addition procedure.
10. General amine addition procedure using solution M5.

After step 10, the resin washed with DMF (5 x 5 mL), dichloromethane (10 x 5 mL), and then dried under high vacuum. This resin was called **PR-1**. Next, a very small aliquot of resin was cleaved using 95% TFA/H<sub>2</sub>O for 1 h at room temperature. The solution was diluted with water, filtered, and checked by HPLC and ESI-MS. HPLC showed a single peak for product and the ESI-MS results are listed below.

Expected mass of product = 722.41

Found [M+H]<sup>+</sup> = 723.3, [M+Na]<sup>+</sup> = 745.4

#### *D.2.5 Procedure for first mini-PEG coupling (Synthesis of Resin **PR-2**)*

The resin, **PR-1**, from the peptoid synthesis procedure above was swelled in DMF for 30 min and drained. A separate vial was charged with 180 mg of Fmoc-8-amino-3,6-dioxaoctanoic acid (Fmoc-AEEA, **8**, CAS 166108-71-0, Peptides International), 244 mg of PyBOP, 92  $\mu$ L of DIEA, and 1 mL of DMF. This reaction mixture was stirred at 23 C for 15 min and then added to the swelled resin **PR-1**. The reaction mixture was then put in a shaker at 37 C for 8 h. The resin washed with DMF (5 x 5 mL), dichloromethane (10 x 5 mL), and then dried under high vacuum to give resin **PR-2** (Yield = 351 mg of dry resin). Next, a very small aliquot of resin was cleaved using 95% TFA/H<sub>2</sub>O for 1 h at room temperature. The solution was diluted with water, filtered, and checked by HPLC and ESI-MS. HPLC showed a single peak for product and the ESI-MS results are listed below.

Expected mass of product = 1089.55

Found  $[M+H]^+ = 1090.4$ ,  $[M+Na]^+ = 1112.5$

#### D.2.6 Preparation of peptoid 5-FAM (1)

The resin **PR-2** (20 mg) from the peptoid synthesis procedure above was swelled in DMF for 20 min and drained. Next, the Fmoc protecting group was removed by addition of 20% piperidine in DMF (2 x 4 mL for 10 min each). The resin was then washed with DMF (10 x 2 mL) then 5 mL of anhydrous DMF. Next, the deprotected resin was treated with a solution of 2 mg of 5-carboxyfluorescein succinimidyl ester (5-FAM-SE, Molecular Probes, C2210) in 300  $\mu$ L of DMF. The reaction flask was covered with foil and agitated at 23 C for 1 h followed by 37 C for 1 hr. Next, the resin was drained, washed with DMF (10 x 1 mL), dichloromethane (10 x 1 mL), and dried under high vacuum. Next, the FAM labeled peptoid was cleaved from resin using 95% TFA/2.5% H<sub>2</sub>O/2.5% triisopropylsilane for 1 h. The crude product was purified by preparative reverse phase HPLC eluting from 10% acetonitrile/90% (0.1% TFA-H<sub>2</sub>O) to 50% acetonitrile/50% (0.1% TFA-H<sub>2</sub>O) over 70 min. The pure product eluted at minute 47 to give 5-FAM (1.74  $\mu$ mol). Extinction coefficient for 5-FAM is 75000 at pH = 9.0,  $\lambda_{\max} = 498$  nm.

**1** (5-FAM): ESI-MS calculated for  $[M+H]^+$ : 1226.5, observed  $[M+H]^+$ : 1226.4

#### D.2.7 Preparation of peptoid 6-FAM (2)

Peptoid 6-FAM (**2**) was synthesized using the exact same procedure as for making 5-FAM except 6-carboxyfluorescein succinimidyl ester (6-FAM-SE, Molecular Probes, C-6164) was used. Purification was performed in the same manner and the elution time for the pure product was the same to give 6-FAM (1.13  $\mu$ mol). Extinction coefficient for 6-FAM is 75000 at pH = 9.0,  $\lambda_{\max} = 498$  nm.

**2** (6-FAM): ESI-MS calculated for  $[M+H]^+$ : 1226.2, observed  $[M+H]^+$ : 1226.2

### D.3 Notes and References

1. Appella, D. H., Christianson, L. A., Klein, D. A., Powell, D. R., Huang, X., Barchi, J. J., Gellman, S. H. Residue-based control of helix shape in beta-peptide oligomers. *Nature* **1997**, 387, 381–384.
2. Seebach, D., Schaeffer, L., Brenner, M. & Hoyer, D. Design and Synthesis of gamma-Dipeptide Derivatives with Submicromolar Affinities for Human Somatostatin Receptors. *Angew. Chem. Int. Ed.* **2003**, 42, 776–778.
3. Zuckermann, R. N., Kerr, J. M., Kent, S. B. H., Moos, W. H. Efficient method for the preparation of peptoids [oligo (N-substituted glycines)] by submonomer solid-phase synthesis. *J. Am. Chem. Soc.* **1992**, 114, 10646–10647.

4. Yu, P., Liu, B., Kodadek, T. A high-throughput assay for assessing the cell permeability of combinatorial libraries. *Nat Biotechnol* **2005**, *23*, 746–751.
5. Xiao, X., Yu, P., Lim, H. S., Sikder, D., Kodadek, T. A Cell-Permeable Synthetic Transcription Factor Mimic. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865–2868.

## **Appendix E: Cell Uptake Studies of 4G Targeting Polyamides**

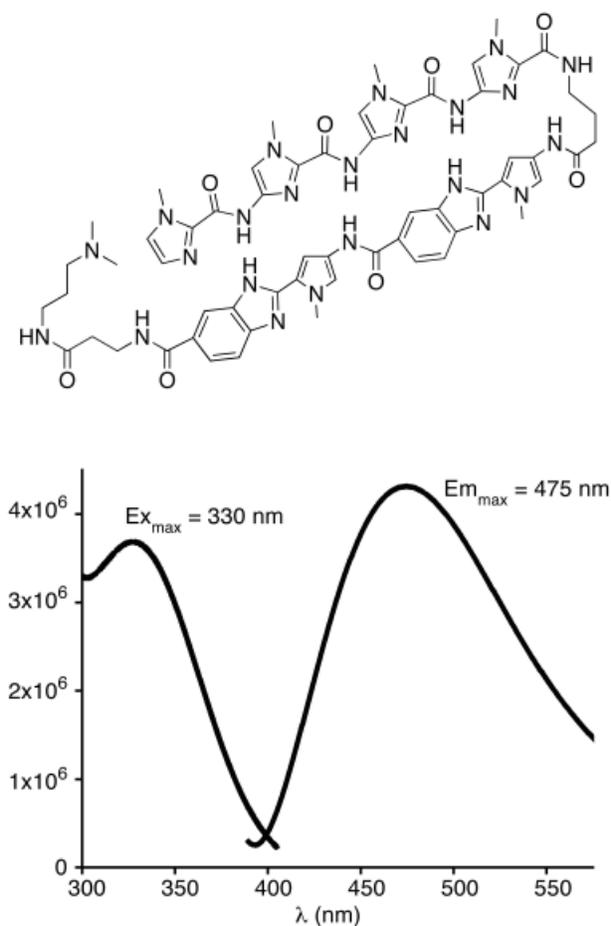
*The research in this appendix is an extension of results from a combination of Chapter 7 and 8. Julie A. Poposki (Caltech) is thanked for HeLa cell culture work and assistance with cell uptake experiments.*

**Abstract**

A benzimidazole-containing polyamide targeted to the sequence 5'-WGGGGW-3' was discovered to have fluorescent properties in aqueous solution in the presence and absence of DNA. This set of properties allowed for the unique opportunity to interrogate polyamide cell uptake in the absence of fluorophore conjugation, which alters the molecular architecture of a compound and often influences cell permeability. A 23-member library of the core benzimidazole polyamide with variations at the tail and turn was synthesized and uptake properties were evaluated in HeLa cells using 2-photon confocal laser scanning microscopy. Investigations of cellular localization using organelle specific dyes were also conducted along with the elucidation of an efflux mechanism for certain polyamides in the library.

## E.1 Introduction

The core benzimidazole containing polyamide **2** from Chapter 6<sup>1</sup> (Figure E.1) was discovered to possess a unique set of photophysical properties where the emission maximum was found to be 475 nm and the excitation maximum was  $\lambda = 330$  nm. Excitation in the range of 380–390 nm was used to minimize photobleaching while still producing a detectable amount of emission at 500 nm. This inherent fluorescent property allowed for the unique opportunity to study cellular localization of this polyamide architecture without attachment of a fluorophore, which can often influence the cellular uptake process.<sup>2-4</sup> A library of polyamides (Figure E.2) containing the fluorescent core was synthesized, where the functionality at the turn and tail was modified in an effort to alter the overall uptake properties. HeLa cells were used for all studies and the polyamides were imaged using two-photon confocal laser scanning microscopy. After assessing the cellular uptake properties of the library of polyamide tail and turn modifications a study was conducted to determine the location of the polyamides in cells (mitochondria, nuclei, vesicles, lysosomes, etc.). Additionally, a polyamide that was found to remain extracellular was investigated to discriminate between cellular impermeability versus efflux using the racemic efflux pump inhibitor ( $\pm$ )-Verapamil (2-(3,4-dimethoxyphenyl)-5-[2-(3,4-dimethoxyphenyl)ethylmethyl-amino]-2-(1-methylethyl)pentanenitrile, CAS 52-53-9). The effect of ethanol fixing on the cellular uptake of polyamides as a function of time was also investigated and the ability of common transfection reagents to influence polyamide uptake was studied.

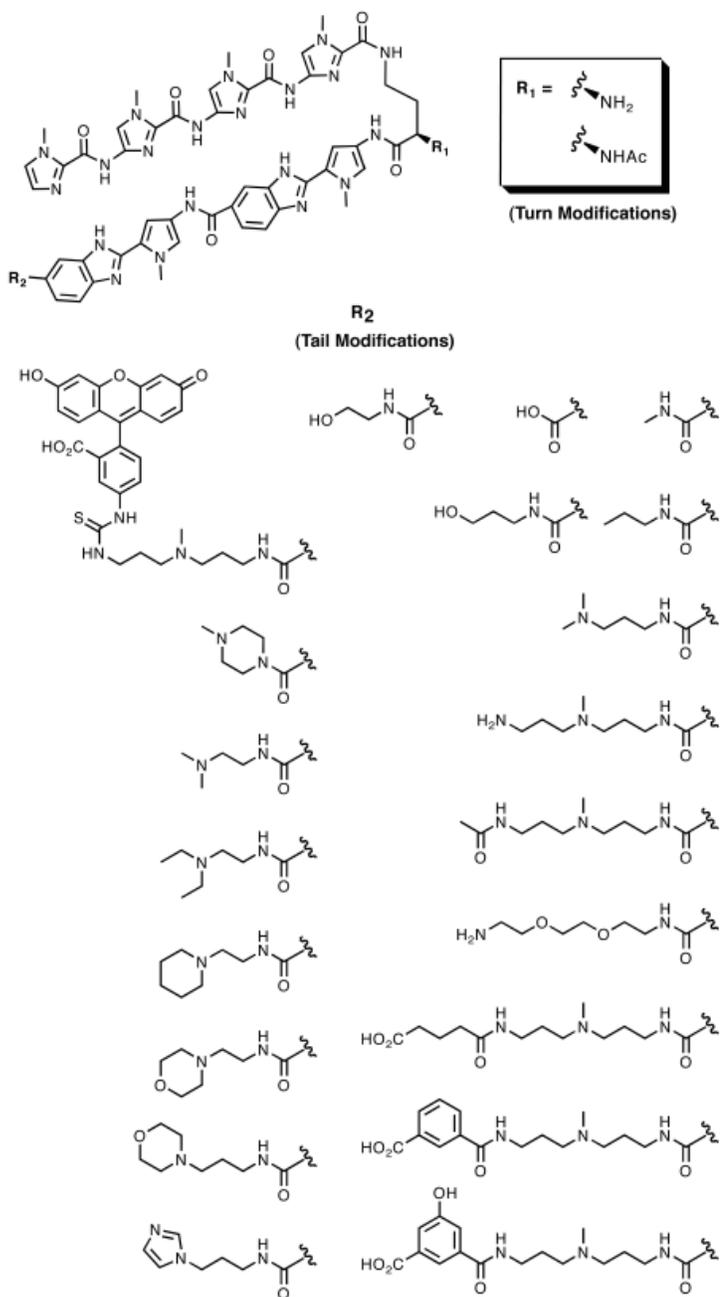


**Figure E.1** Fluorescent polyamide **2** from Chapter 7 of this thesis. (bottom) Overlaid emission and excitation spectra of polyamide **2**.

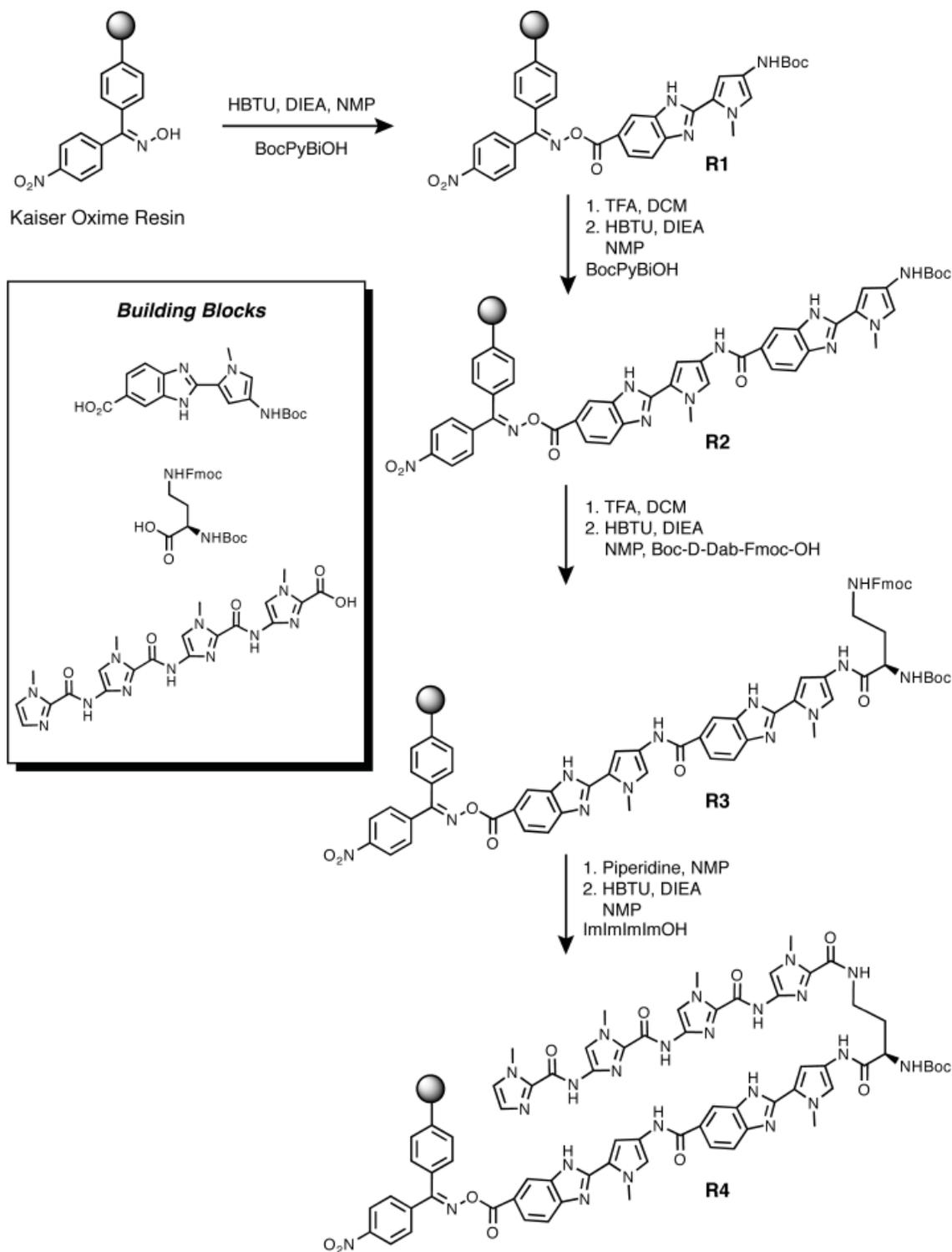
## E.2 Results and Discussion

A focused library of 23 compounds was synthesized containing tail and turn modifications on a common polyamide core (Figure E.2). The polyamide was synthesized on Kaiser Oxime resin by standard solid-phase synthesis methods as shown in Scheme E.1 to provide a base resin loaded with the desired polyamide core used for tail and turn diversification. With the core in hand, diversification commenced with deprotection and cleavage from resin to afford triamine derivatives **3**, **9**, **20**, **22**, and **23** (Scheme E.2). Deprotection and resin cleavage with a small set of amine nucleophiles produced compounds **1**, **2**, **4-8**, **10**, **11**, **13-16**, and **19**. Acylation prior to deprotection followed by cleavage from resin with amine nucleophiles produced compounds **12**, **17**, and **18**. Deprotection and cleavage from resin with lithium hydroxide produce compound **21**.

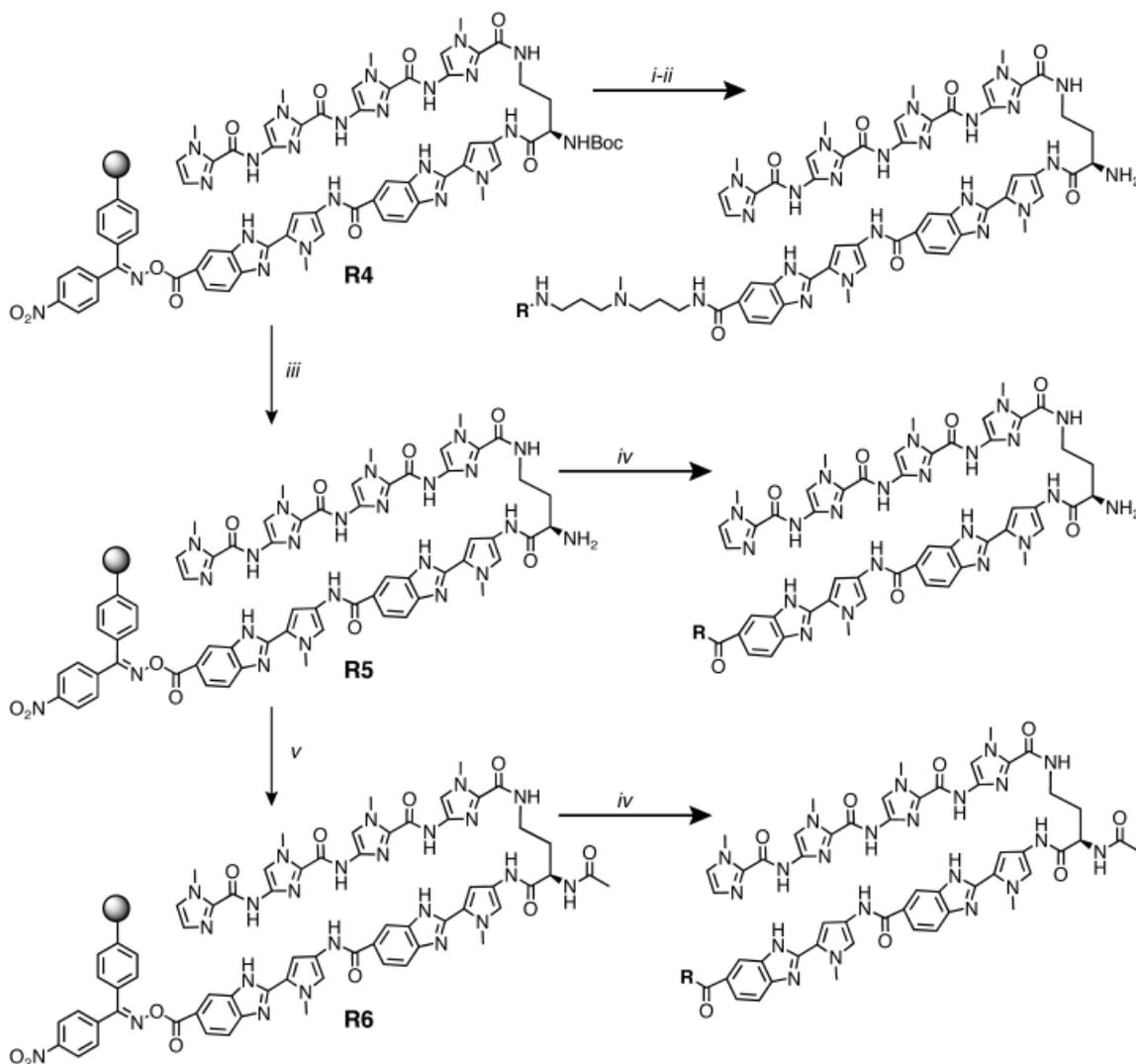
The cell permeability of this small focused library of compounds was evaluated in HeLa cells using 2-photon laser-scanning confocal microscopy and the results are shown in Figures E.3–E.8. The results are grouped according to their cellular localization profile with Figures E.3–E.6 showing compounds that appear to be cell permeable but non-nuclear,



**Figure E.2** Tail and turn modifications used for the construction of a 23 compound benzimidazole-containing polyamide library.



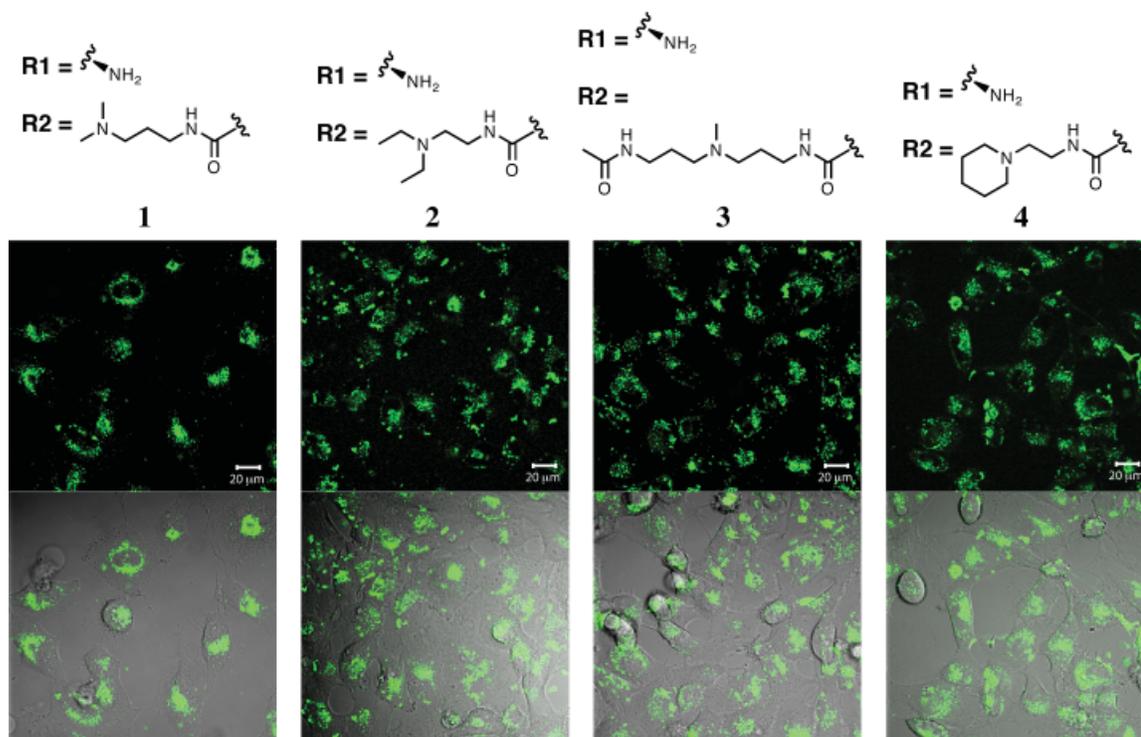
**Scheme E.1.** Synthesis of **R4**. Standard solid-phase synthesis methods from Chapter 7<sup>1</sup> were used.



**Scheme E.2.** Synthesis of compound library. Reaction conditions: (i) NMP, Triamine, 80 °C (Heat Block), 15-20 min; (ii) Conjugation to triamine, TFA, DCM; (iii) TFA, DMC; (iv) NMP, Nucleophile, 80 °C (Heat Block), 15-20 min; (v) Ac<sub>2</sub>O, NMP.

producing punctate cytoplasmic staining pattern. The compounds shown in Figure E.7 appear to be primarily extracellular and aggregated or precipitated with the first two compounds coating the exterior cell membranes. Figure E.8 shows compounds that are primarily extracellular although in a few cases there is a small amount of non-nuclear cellular localization observed.

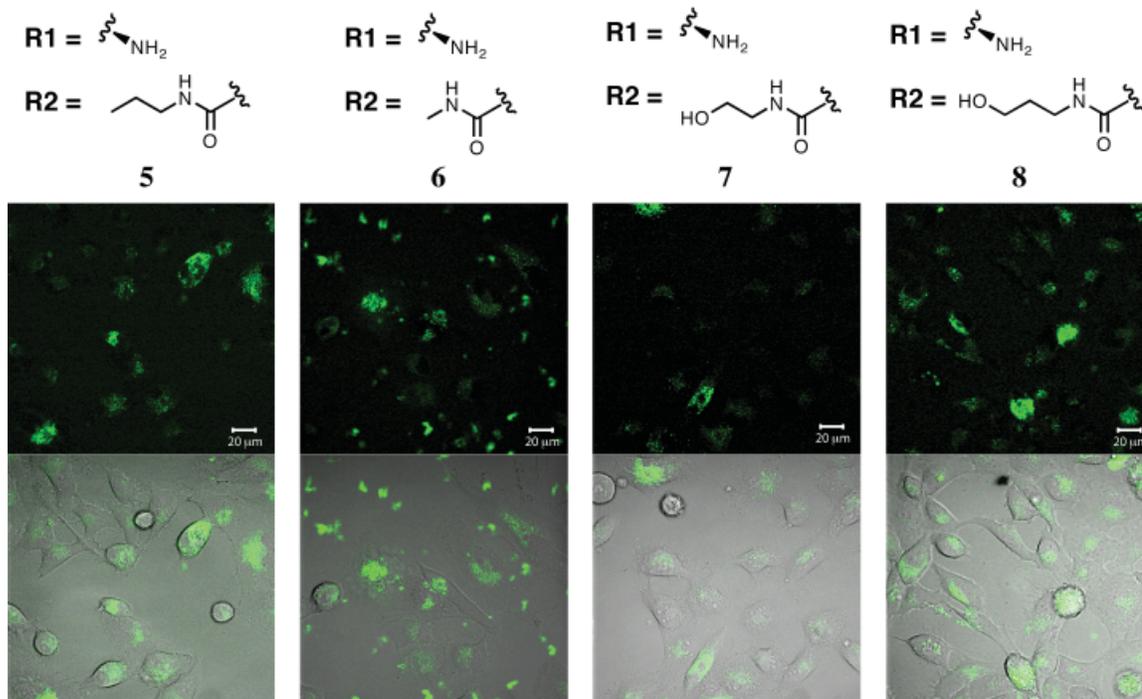
In an effort to determine the location of the compounds inside the cells in Figures E.3–E.6 (punctate cytoplasmic staining) two organelle specific dyes were utilized. MitoTracker Red (CM-H2XRos) was used to assess the possibility of mitochondrial localization and LysoTracker Red (DND-99, a lysosome and trans-golgi stain) to assess localization in acidic lysosomes within the cell. Previous reports by Lown and coworkers demonstrated that a distamycin-fluorophore conjugate



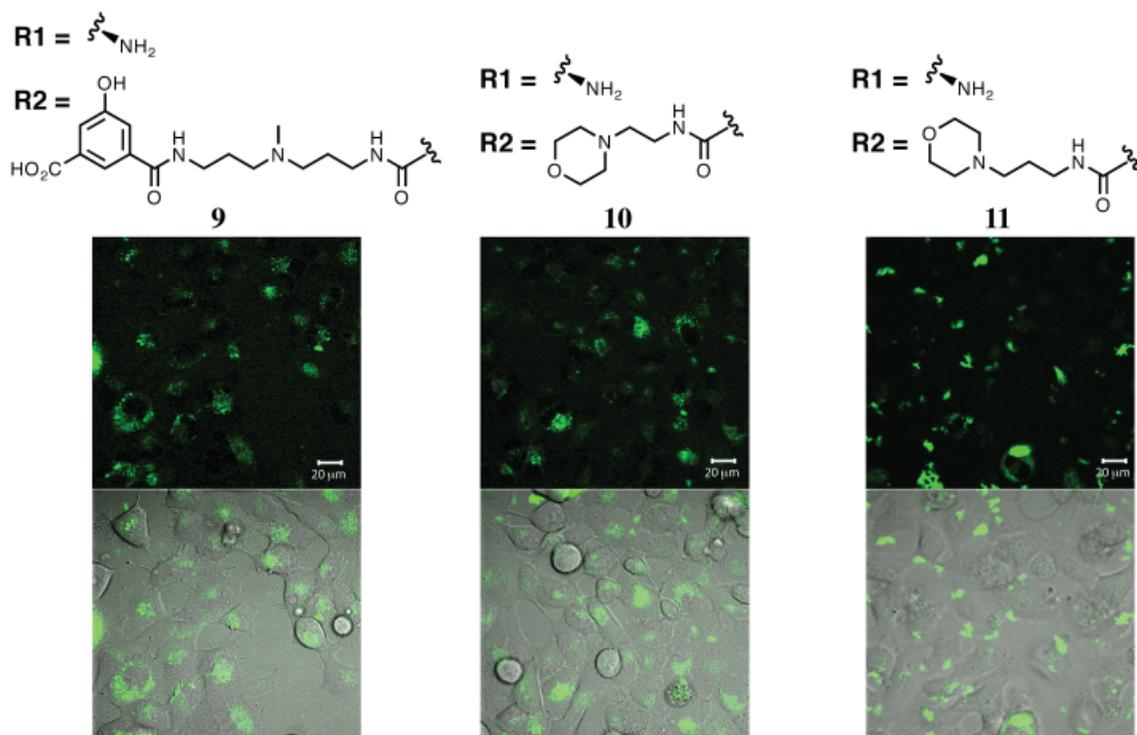
**Figure E.3** Fluorescent polyamide cell uptake studies for compounds **1-4**. 2-Photon Laser Imaging Setup: Coherent Chameleon 2-photon Laser, ( $\lambda = 810$  nm, 5% power), HFT KP 680 dichroic, BP 480-520 nm HeLa cells treated with 2mM **1-4** and imaged using 2-photon laser microscopy:

primarily localizes in the mitochondria of human ovarian adenocarcinoma cells.<sup>5</sup> The mechanism of mitochondrial staining by MitoTracker relies upon passive diffusion of the pro-fluorescent dye into the cell where it is oxidized and sequestered in the mitochondria (Figure E.9). Once in the mitochondria, nucleophilic attack resulting in thiol conjugation by proteins and peptides serves to retain the dye. Polyamide treated HeLa cells were dosed with MitoTracker prior to dual imaging using confocal laser scanning microscopy. The results show that the mitochondria appear as a diffuse cytoplasmic stain (Figure E.10, upper left live cell panel, magenta). The upper right panel shows the polyamide imaged with an orthogonal wavelength to Mitotracker and the bottom right panel shows that the two do not colocalize. It appears that the polyamide is not localized in mitochondria. Interestingly, ethanol fixing (fixed cell panel, bottom right, Figure E.10) of the cells leads to rapid nuclear trafficking of the polyamide however the diffuse mitochondria staining appears unchanged. It appears that unlike the Distamycin conjugates studied by Lown this class of polyamides does not localize in the mitochondria. However, this is not a surprising result given the dramatic differences in molecular structure and cell type.

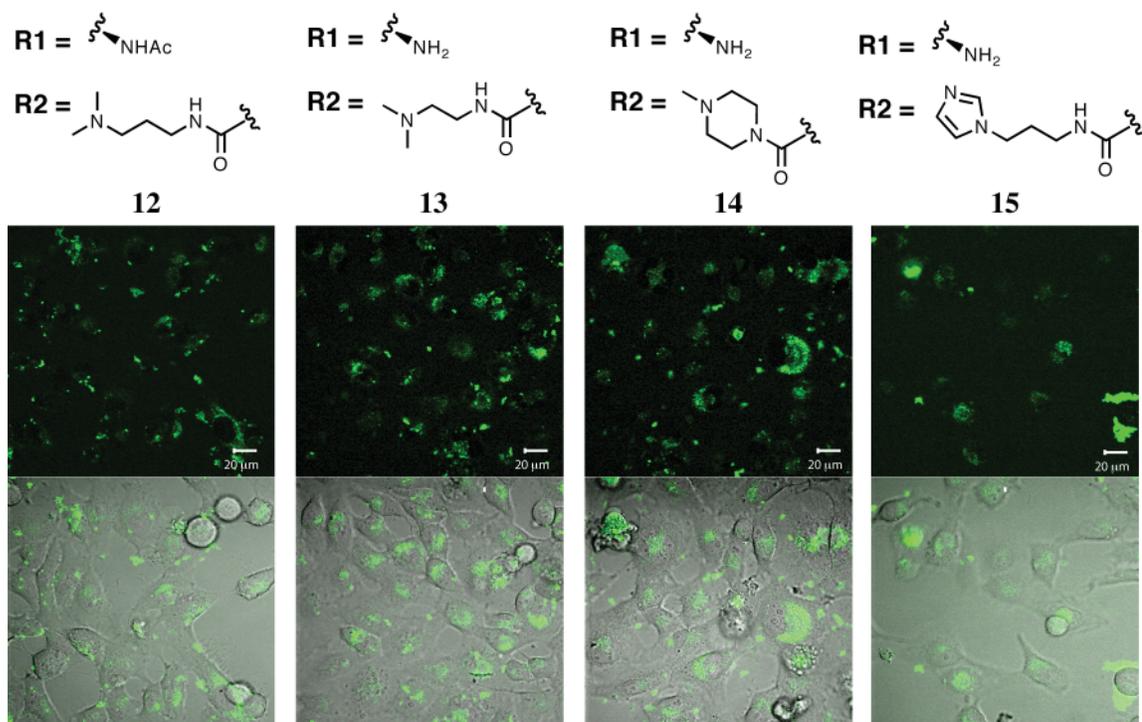
Next, LysoTracker was used to determine if the polyamides were localized in acidic lysosomes of the cells. Figure E.11 shows the results of this study and it appears that the fluorescence



**Figure E.4** Fluorescent polyamide cell uptake studies 5-8. 2-Photon laser imaging setup and polyamide dosing concentration as described in Figure E.2.



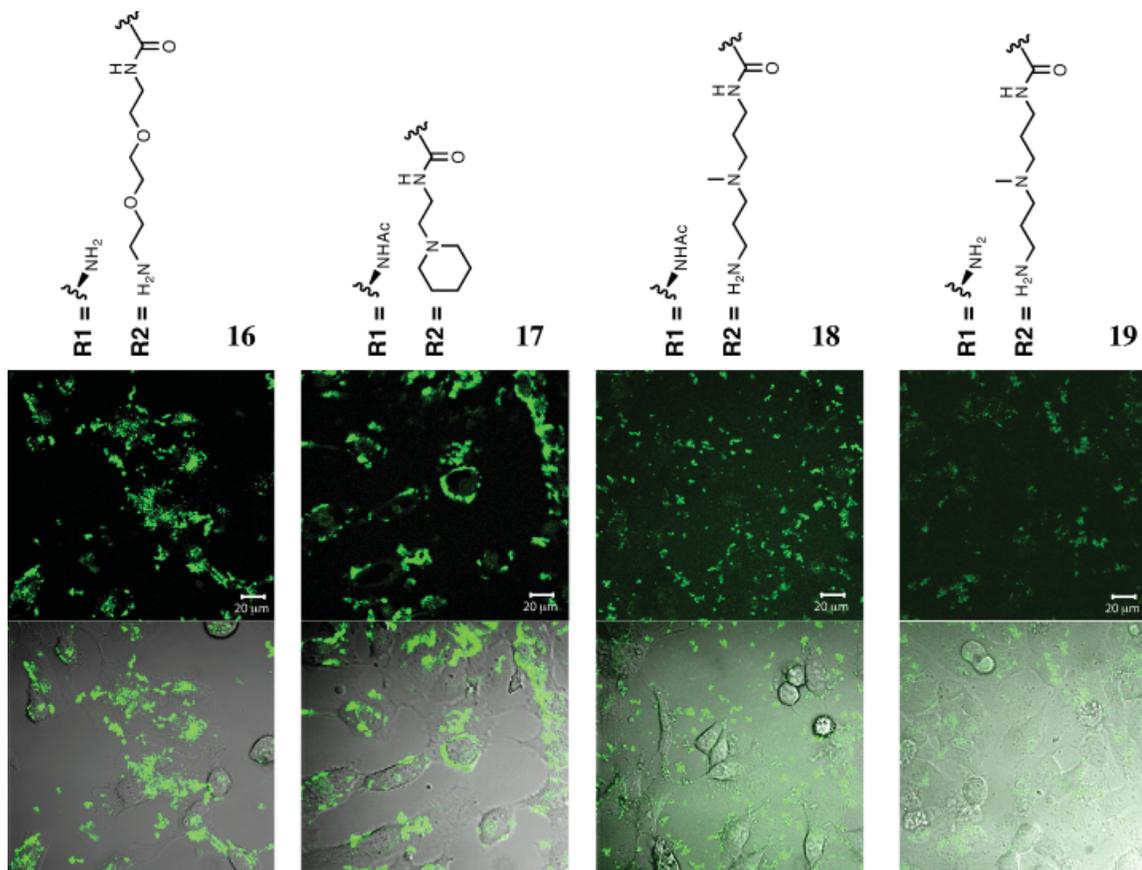
**Figure E.5** Fluorescent polyamide cell uptake studies 9-11. 2-Photon laser imaging setup and polyamide dosing concentration as described in Figure E.2.



**Figure E.6** Fluorescent polyamide cell uptake studies **12-15**. 2-Photon laser imaging setup and polyamide dosing concentration as described in Figure E.2.

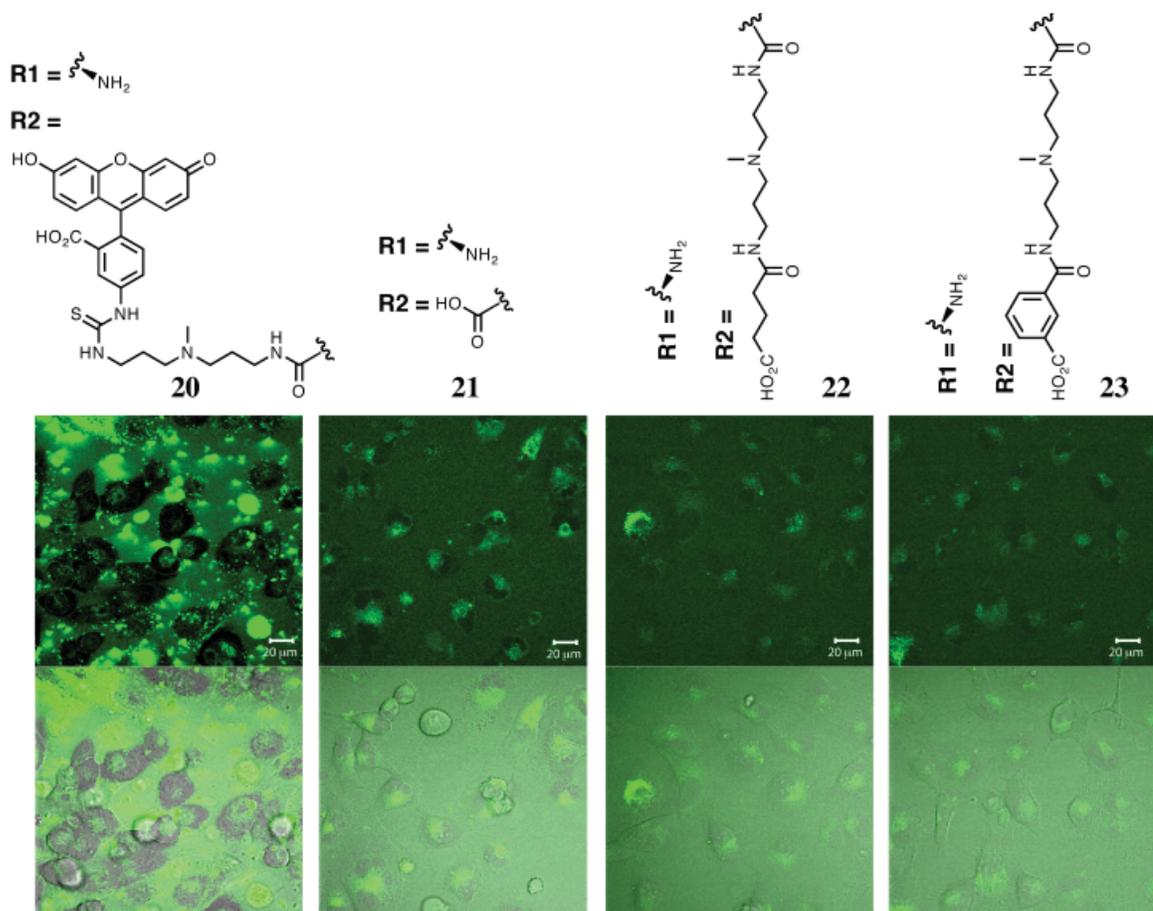
signal of the internalized polyamides colocalizes with the fluorescence signal of LysoTracker, indicating that the polyamides are likely sequestered in acidic vesicles within the cells. This result is consistent with previous studies showing that in all cases 8-ring polyamides with 4 imidazoles show poor nuclear uptake profiles.<sup>3</sup> Weakly basic amines have also been shown to selectively accumulate within low pH cellular compartments such as lysosomes.<sup>6</sup> It should also be noted that the *N,N*-dimethylaminopropylamine tail commonly used to terminate polyamides and other classes of minor groove binders is very similar to the functionality present in LysoTracker responsible for targeting it to the lysosomes and acidic vesicles of cells. Additionally, high imidazole content has been correlated with low cell uptake and the four contiguous imidazoles in the polyamide core may be basic enough to target the polyamides to the lysosomes without the influence of any other functionality, however benzimidazole involvement can not be ruled out.<sup>3,4</sup>

The uptake properties of polyamide **23** in Figure E.8 were investigated in an effort to determine their mechanism of exclusion from the cells. Polyamide **23** (Figure E.8) dosed HeLa cells were treated with ( $\pm$ )-Verapamil and imaged using 2-photon laser confocal microscopy. The results in Figure E.12 show that treatment of polyamide dosed cells with ( $\pm$ )-Verapamil causes uptake

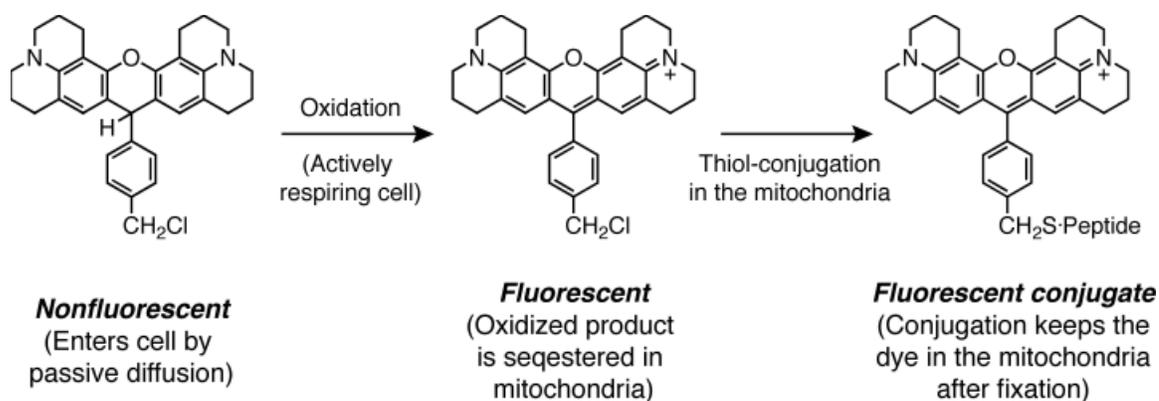


**Figure E.7** Fluorescent polyamide cell uptake studies **16-19**. 2-Photon laser imaging setup and polyamide dosing concentration as described in Figure E.2.

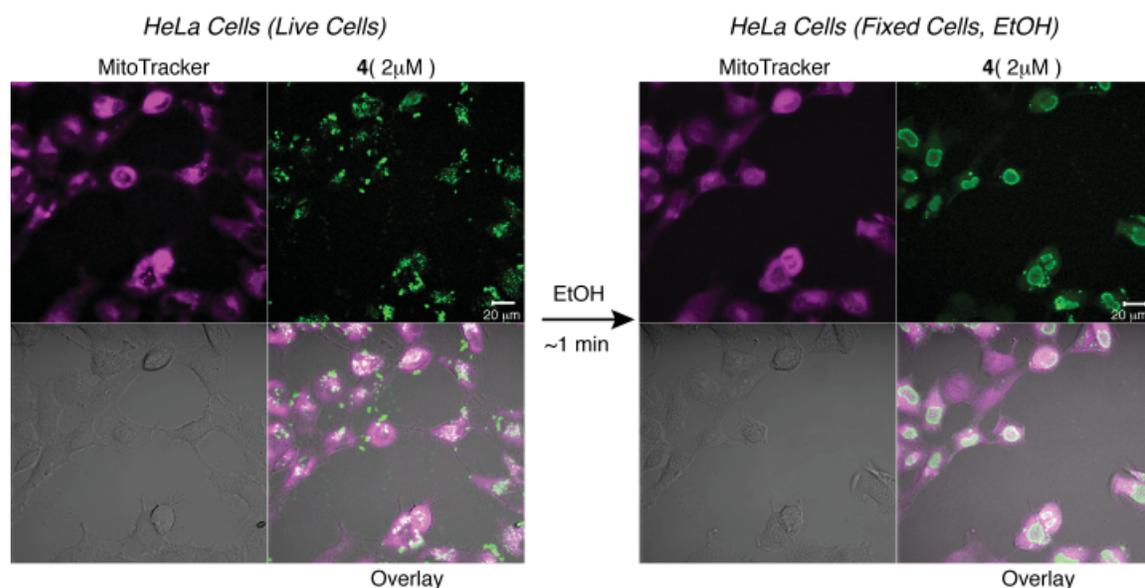
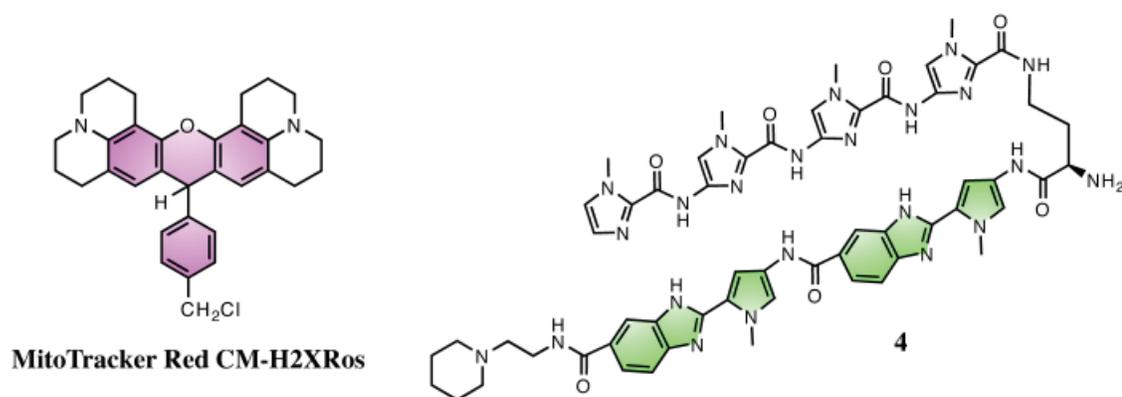
of the polyamide however the polyamide appears to be localized in acidic vesicles (potentially lysosomes) as in Figures E.3–E.6. Figure E.12 also shows that treatment of the cells with ethanol allows the polyamide to readily traffic to the nucleus with significant nuclear localization after 2 h and complete localization after 4 h. Next the effect of a commonly used transfection reagent (Lipofectamine 2000) on polyamide nuclear localization was investigated (Figure E.13). The results show that the transfection reagent has little to no effect on the nuclear localization of this class of polyamide, however increased extracellular aggregation was observed.



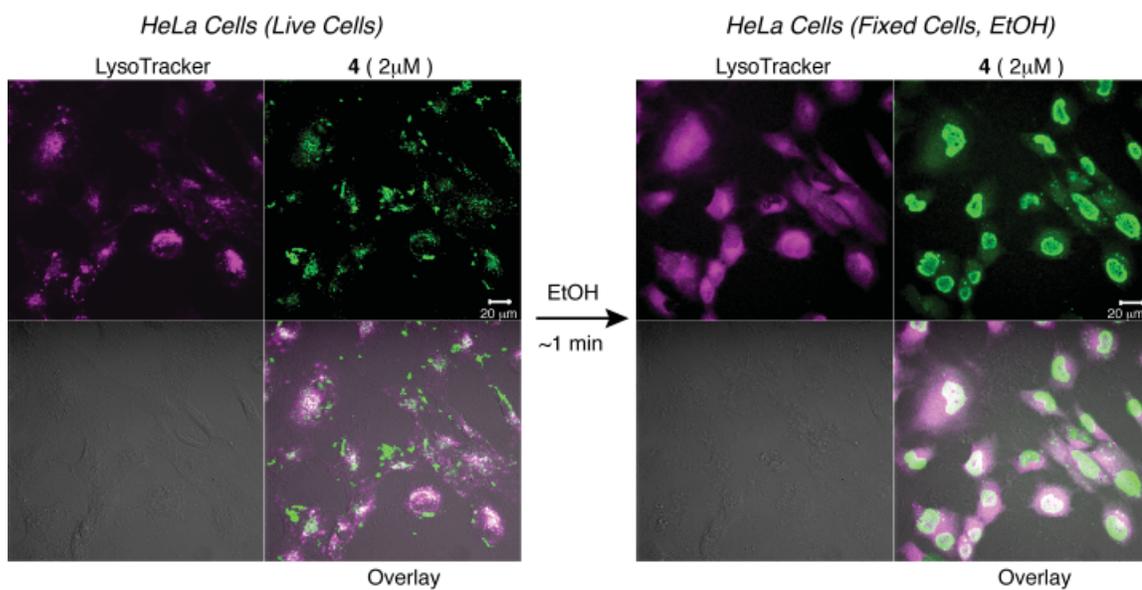
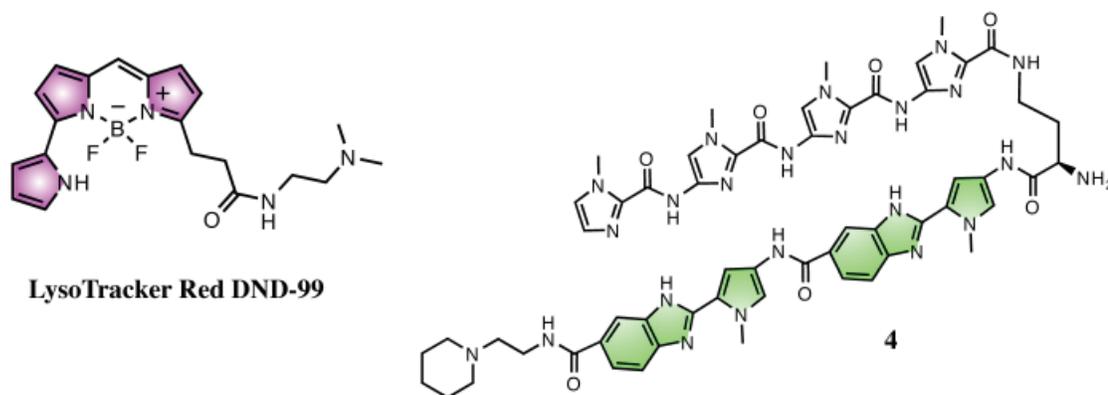
**Figure E.8** Fluorescent polyamide cell uptake studies **20–23**. 2-Photon laser imaging setup and polyamide dosing concentration as described in Figure E.2.



**Figure E.9** Mechanism of mitochondrial staining using MitoTracker.

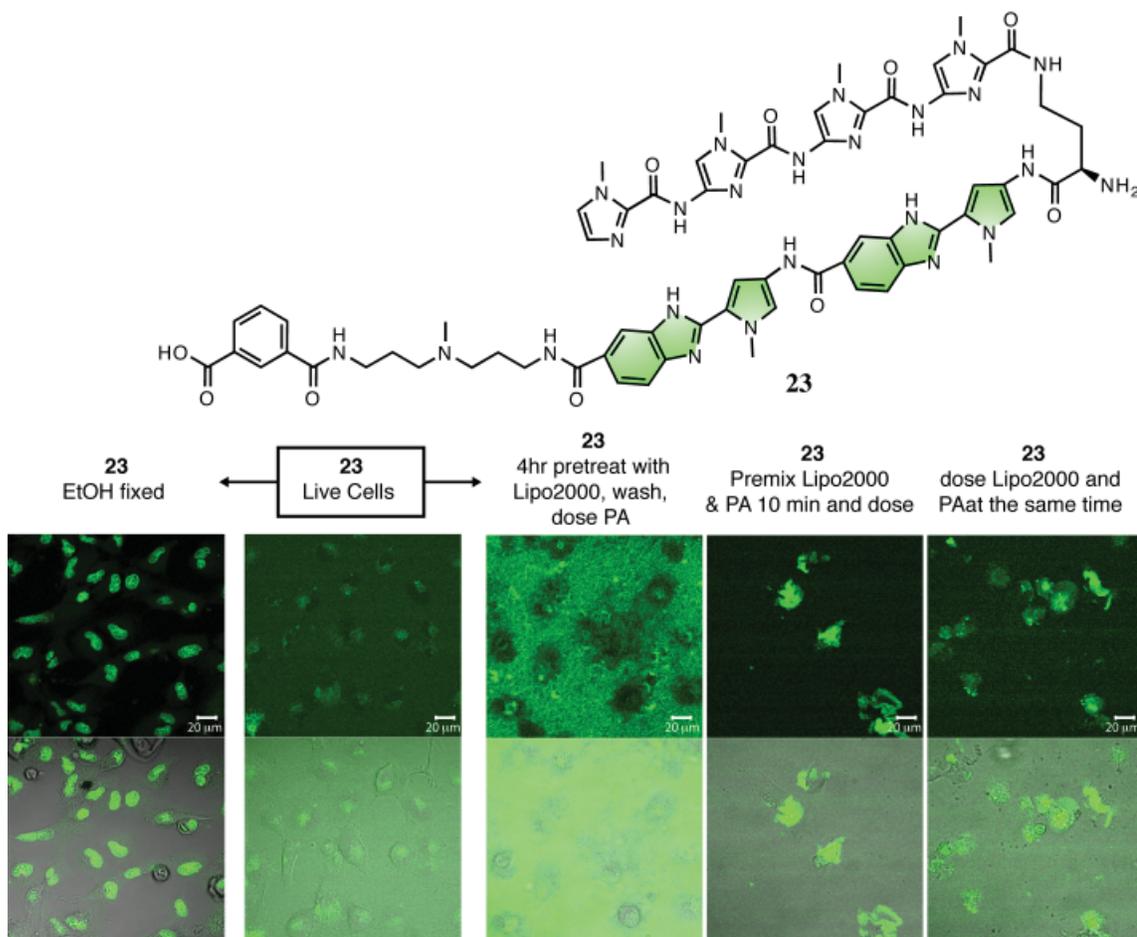


**Figure E.10** Results of HeLa cell uptake study using MitoTracker and compound **4**. Dual Laser Imaging Setup: Coherent Chameleon 2-photon Laser, ( $\lambda = 810$  nm, 5% power), HFT KP 680 dichroic, BP 480–520 nm MitoTracker, HeNe Laser ( $\lambda = 543$  nm, 10% power), HFT KP 680 dichroic, BP 565–615 nm



**Figure E.11** Results of HeLa cell uptake study using LysoTracker and compound **4**. Dual Laser Imaging Setup: Coherent Chameleon 2-photon Laser, ( $\lambda = 810$  nm, 5% power), HFT KP 680 dichroic, BP 480-520 nm LysoTracker, HeNe Laser ( $\lambda = 543$  nm, 10% power), HFT KP 680 dichroic, BP 565-615 nm





**Figure E.13** Cell uptake study with Lipofectamine 2000 and compound **23** (concentration =  $2\mu\text{M}$  in all experiments). Lipofectamine 2000 (Lipo2000) concentration =  $1\mu\text{L}/150\mu\text{L}$  media. HeLa cells used in all experiments. 2-Photon Laser Imaging Setup: Coherent Chameleon 2-photon Laser, ( $\lambda = 810\text{ nm}$ , 5% power), HFT KP 680 dichroic, BP 480-520 nm

### E.3 Conclusion

The results presented in these studies of a focused 23-member library of tail and turn modified polyamides demonstrate that the uptake properties of benzimidazole-containing polyamides, targeted to the sequence 5'-WGGGG-3', can be divided into 2 broad categories. The first category includes compounds that are intracellular but non-nuclear, displaying a punctate cytoplasmic staining. The second category of compounds remain extracellular and have overall poor cellular uptake properties. Localization studies of non-nuclear intracellular polyamides indicated lysosomal sequestering as determined by LysoTracker Red DND-99 colocalization studies. This is in stark contrast to observations by Lown and coworkers who found that fluorescently-labeled distamycin derivatives localized primarily to the mitochondria of human ovarian adenocarcinoma

cells (SKOV-3).<sup>5</sup> In addition, the mechanism of poor uptake for compound **23** was investigated using the efflux pump inhibitor ( $\pm$ )-verapamil<sup>7</sup> and results from this study indicate that inhibition of efflux pumps causes rapid cellular uptake, however lysosomal sequestration is observed as indicated by colocalization studies with LysoTracker. As documented previously, ethanol fixing of cells causes rapid nuclear uptake due to increased membrane permeability.<sup>2-4</sup> It has been reported that weakly basic amines selectively accumulate in cellular compartments with low internal pH and can be used to investigate the biosynthesis and pathogenesis of lysosomes.<sup>6</sup> In addition to the use of weakly basic amines on the tail and turns of polyamides, high imidazole content has also been shown to be a negative determinant of nuclear localization.<sup>3,4</sup> Currently, the cell uptake of high imidazole-containing polyamide sequences<sup>1</sup> remains an unsolved problem.

#### E.4 Experimental

Synthesis of polyamides **1–23** was performed according to general solid phase synthesis protocols outlined in Chapter 7 and MALDI-TOF-MS data for all compounds is presented in Section E.6. The setup for 2-photon confocal laser scanning microscopy is reported in figure captions in this appendix.

#### E.5 Notes and References

1. Chenoweth, D. M., Poposki, J. A., Marques, M. A., and Dervan, P. B. Programmable oligomers targeting 5'-GGGG-3' in the minor groove of DNA and NF-kappaB binding inhibition. *Bioorg. Med. Chem.* **2007**, *15*, 759–770.
2. Belitsky, J. M., Leslie, S. J., Arora, P. S., Beerman, T. A., and Dervan, P. B. Cellular uptake of *N*-methylpyrrole/*N*-methylimidazole polyamide-dye conjugates. *Bioorg. Med. Chem.* **2002**, *10*, 3313–3318.
3. Best, T. P., Edelson, B. S., Nickols, N. G., and Dervan, P. B. Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12063–12068.
4. Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., and Dervan, P. B. Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res.* **2004**, *32*, 2802–2818.
5. Sharma, S. K., Morrissey, A. T., Miller, G. G., Gmeiner, W. H., and Lown, J. W. Design, synthesis, and intracellular localization of a fluorescently labeled DNA binding polyamide related to the antibiotic distamycin. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 769–772.
6. Griffiths, G., Hoflack, B., Simons, K., Mellman, I., and Kornfeld, S. The mannose 6-phosphate receptor and the biogenesis of lysosomes. *Cell.* **1988**, *52*, 329–341.
7. Bellamy, W. T. P-glycoproteins and multidrug resistance. *Annu. Rev. Pharmacol. Toxicol.* **1996**, *36*, 161–183.

## E.6 Spectra and Supplemental Information

- 1: MALDI-TOF-MS calculated for [M]: 1155.51, observed [M+H]<sup>+</sup>: 1156.13
- 2: MALDI-TOF-MS calculated for [M]: 1348.63, observed [M+H]<sup>+</sup>: 1348.54
- 3: MALDI-TOF-MS calculated for [M]: 1240.57, observed [M+H]<sup>+</sup>: 1241.63
- 4: MALDI-TOF-MS calculated for [M]: 1181.53, observed [M+H]<sup>+</sup>: 1182.50
- 5: MALDI-TOF-MS calculated for [M]: 1112.47, observed [M+H]<sup>+</sup>: 1113.76
- 6: MALDI-TOF-MS calculated for [M]: 1084.44, observed [M+H]<sup>+</sup>: 1085.58
- 7: MALDI-TOF-MS calculated for [M]: 1114.45, observed [M+H]<sup>+</sup>: 1115.53
- 8: MALDI-TOF-MS calculated for [M]: 1128.47, observed [M+H]<sup>+</sup>: 1129.78
- 9: MALDI-TOF-MS calculated for [M]: 1362.57, observed [M+H]<sup>+</sup>: 1363.71
- 10: MALDI-TOF-MS calculated for [M]: 1183.51, observed [M+H]<sup>+</sup>: 1184.55
- 11: MALDI-TOF-MS calculated for [M]: 1197.52, observed [M+H]<sup>+</sup>: 1198.74
- 12: MALDI-TOF-MS calculated for [M]: 1197.52, observed [M+H]<sup>+</sup>: 1198.45
- 13: MALDI-TOF-MS calculated for [M]: 1141.50, observed [M+H]<sup>+</sup>: 1142.37
- 14: MALDI-TOF-MS calculated for [M]: 1153.50, observed [M+H]<sup>+</sup>: 1154.61
- 15: MALDI-TOF-MS calculated for [M]: 1178.49, observed [M+H]<sup>+</sup>: 1179.76
- 16: MALDI-TOF-MS calculated for [M]: 1201.52, observed [M+H]<sup>+</sup>: 1202.90
- 17: MALDI-TOF-MS calculated for [M]: 1223.54, observed [M+H]<sup>+</sup>: 1224.59
- 18: MALDI-TOF-MS calculated for [M]: 1240.57, observed [M+H]<sup>+</sup>: 1241.60
- 19: MALDI-TOF-MS calculated for [M]: 1198.55, observed [M+H]<sup>+</sup>: 1199.50
- 20: MALDI-TOF-MS calculated for [M]: 1587.59, observed [M-FITC]<sup>+</sup>: 1199.90;  
ESI-MS calculated for [M]: 1587.6, observed [M+H]<sup>+</sup>: 1588.1
- 21: MALDI-TOF-MS calculated for [M]: 1071.41, observed [M+H]<sup>+</sup>: 1072.51
- 22: MALDI-TOF-MS calculated for [M]: 1312.58, observed [M+H]<sup>+</sup>: 1313.53
- 23: MALDI-TOF-MS calculated for [M]: 1346.57, observed [M+H]<sup>+</sup>: 1347.78