

# **DNA Binding Polyamides in Biological Systems**

Thesis by  
Jason M. Belitsky

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

California Institute of Technology  
Pasadena, California  
2002

(Defended May 21, 2002)

© 2002

Jason M. Belitsky

All Rights Reserved



*To my friend and lab-mate*  
*Brock Safronoff*  
*and other victims of the 9/11 tragedy*

## Acknowledgments

First and foremost I want to thank my family for all their support. I'm particularly glad that my brother Darren moved out here to experience LA with me these last two years. The strength and togetherness that my parents have shown in overcoming their health problems in recent years will always be an inspiration to me. I love them very much and am extremely grateful that we were able to reach this milestone together in good health.

Next, I want to thank my advisor, Peter Dervan, for his enthusiasm, encouragement, and for providing me with good research projects and even better co-workers. I also want to thank the members of my committee, Rich Roberts, Doug Rees, and Bob Grubbs, for all their support. I am grateful to the professors whose classes I had the opportunity to attend, especially to Andy Myers, Barbara Imperiali, and Rich Roberts, and to the professors that I had the opportunity to TA for, particularly Peter Dervan, Bob Grubbs, Dennis Dougherty, and Ken Williamson. I would also like to thank Harry Gray for the weekly Organometallics seminar, where I able to see a great deal of interesting chemistry. The student-run Organic Chemistry Journal Club was also a great learning experience for me, for which I especially have to thank J.P. Morgan, Isaac Carrico, John Stellwagon, Arnab Chatterjee, and Tae-Lim Choi. I would never have been in a position to learn anything at Caltech if not for the great education I received from the Amherst College chemistry department, particularly from my undergraduate thesis advisor David Hansen.

I was very fortunate to have excellent collaborators for my research projects, particularly Stephanie Leslie and Professor Terry Beerman from the Roswell Park Cancer Institute, Fan Yang and Professor Monica Roth from Robert Wood Johnson Medical School, and Professor Joel Gottesfeld from the Scripps Research Institute. In the Dervan lab, I have been extremely fortunate to collaborate with Bobby Arora, Ben Edelson, Doan Nguyen, Nick Wurtz, and Roland Burli. I have very much enjoyed working with Bobby and Ben as well as with Nick Nickols, Leonard Prins, and Amanda Cashin on the uptake

project; growing cells, fighting mold, imaging, and planning our next step. This project would never have gotten off the ground at Caltech without the help of many people (listed in the Acknowledgements in Chapters 4A and 4B), particularly Inder Nangiani of the Caltech Protein Expression Facility.

I'm especially thankful for all the great friends I've made at Caltech, especially Doan Nguyen, Isaac and Liz Carrico, Niki Zacharias, J.P. Morgan, Arnab Chatterjee, Walter Goldberger, Catherine Baker, Girish Aakalu, Mike Farwell, Soojin Kwon, Carlos Basques, Peter Hackley, and Nick Wurtz. "Renaldo" thanks the entire Grubbs group for their friendship, and the Athenaeum for foosball. I want to thank my current house-mates and friends Niki Zacharias, Gabe Brandt, and Sarah Monahan for putting up with me while I was writing my proposals and thesis. I'm grateful to David Malmud, Rob Leibowitz, Scott Mellender, and Steve Lasher for their long and continuing friendship.

Finally I want to thank the entire Dervan group for their friendship, advice, collaboration, and commiseration. I've been surrounded by great scientists in my classmates Doan Nguyen, Nick Wurtz, and Adam Urbach. I'm particularly grateful to have been lab-mates with Doan, Anna Mapp, Paul Floreancig, David Herman, John Trauger, Adam Kerstein, Meredith Howard, Philipp Weyermann, Christoph Briehn, Tim Best, Shira Jacobsen, and John Chevillet. Getting to know and working with Doan, especially, has been a highlight of my Caltech experience. I've also benefited very much from the wisdom and friendship of Ryan Bremer, Clay Wang, Aileen Chang, Christian Melander, Tom Minehan, Ulf Ellervik, and Nick W. I've enjoyed the Amherst connection with Meredith, Adam K., Scott Carter, and Sue Swalley. I've had many interesting conversations with Victor Rucker, and enjoyed getting to know the BI contingent, Bogdan Olenyuk, Pierre Potier, Inger Kers, Konstanze Gottewald, and Simon Friedman. Last but definitely not least, I want to thank three postdocs who helped me out tremendously when I first joined the lab and even after they had left it: Anna Mapp, Paul Floreancig, and Roland Burli. I couldn't ask for better mentors, lab-mates, or friends.

## Abstract

Small molecules that bind selectively to a DNA sequence in the human genome are potentially useful tools for molecular biology and human medicine. Polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) are small molecules that bind DNA according to a set of “pairing rules” with affinities and specificities that rival natural transcription factors. By directly competing with a given transcription factor or other DNA binding protein for its binding site, polyamide can cause inhibition of diverse biological processes, such as retroviral integration and gene transcription. Polyamides are presented which inhibit the *in vitro* integration activities for two retroviruses, M-MuLV and HIV-1. Polyamides are described that inhibit TBP binding to the HER2 promoter, a gene implicated in human breast cancer. Failure to achieve inhibition of HER2 transcription in cell culture led to the surprising discovery that polyamide-fluorescent dye conjugates are cell-permeable, but that nuclear localization does not occur in many cell lines. Efforts toward modified polyamides with enhanced nuclear localization properties are presented. In order to extend the number of sequences amenable to high affinity recognition, alteration of the C-terminal polyamide tail has been investigated. The development of conditions for polyamide solid-phase synthesis on a new resin, which allows for the generation of “truncated tail” polyamides, is presented. During the original route to one of these compounds, an unexpected reaction was uncovered that leads to entirely different C-terminal tails.

## Table of Contents

	page
Acknowledgments.....	iv
Abstract.....	vi
Table of Contents.....	vii
List of Figures and Tables.....	ix
 <b>CHAPTER ONE:</b> Introduction.....	 1
 <b>CHAPTER TWO:</b> Inhibition of Retroviral Integration by DNA Binding Polyamides .....	 19
<b>Chapter 2A:</b> Inhibition of Moloney Murine Leukemia Virus Integration Using Polyamides Targetiing the Long-Terminal Repeat (LTR) Sequences .....	 21
<b>Chapter 2B:</b> Inhibition of HIV-1 Integration by Hairpin Polyamides...	49
 <b>CHAPTER THREE:</b> DNA Binding Polyamides for the Inhibition of HER2 Transcription.....	 62
 <b>CHAPTER FOUR:</b> Cellular Uptake and Localization of DNA Binding Polyamide-Fluorescent Dye Conjugates .....	 89
<b>Chapter 4A:</b> Cellular Uptake of <i>N</i> -Methylpyrrole/ <i>N</i> -Methyl- imidazole Polyamide-Dye Conjugates .....	 91
<b>Chapter 4B:</b> Enhanced Nuclear Localization of Polyamide-Dye Conjugates .....	104

<b>CHAPTER FIVE:</b>	Synthesis and Investigation of DNA Binding	
	Polyamides with New C-Terminal Tails.....	140
<b>Chapter 5A:</b>	Solid-Phase Synthesis of DNA Binding	
	Polyamides on Oxime Resin .....	142
<b>Chapter 5B:</b>	From an Aminooxy $\beta$ -Alanine Analog to Novel	
	Polyamide Tails .....	159

## List of Figures and Tables

<b>CHAPTER ONE</b>	page
Figure 1.1 Structure of double-helical B-form DNA.....	2
Figure 1.2 DNA base pairs and schematic representation of the minor groove .....	3
Figure 1.3 Representative X-ray crystal structures of DNA binding proteins.....	3
Figure 1.4 Structures of compounds which bind in the minor groove .....	4
Figure 1.5 Schematic representation of the polyamide pairing rules.....	5
Figure 1.6 X-ray crystal structure polyamide homodimer.....	6
Figure 1.7 Inhibition of gene transcription by hairpin polyamides.....	8
Figure 1.8 Examples of gene promoters tareted by hairpin polyamides.....	9
Figure 1.9 X-ray crystal structure of the nucleosome core particle.....	12
 <b>CHAPTER TWO</b>	
Figure 2.1 Polyamides designed to bind the M-MuLV LTR.....	26
Figure 2.2 DNase I footprinting experiments.....	30
Figure 2.3 Schematic illustration of <i>in vitro</i> integration assays.....	31
Figure 2.4 Inhibition of 3' processing reactions.....	32
Figure 2.5 Inhibition of strand transfer reactions.....	33
Figure 2.6 Strand transfer into an exogenous substrate .....	35
Figure 2.7 IC <sub>50</sub> of polyamides against M-MuLV integrase.....	36
Figure 2.8 DNA melting experiments .....	42
Figure 2.9 Polyamides designed to bind the HIV-1 U5 LTR.....	50
Figure 2.10 Binding sites for polyamides in the HIV-1 U5 LTR.....	51
Figure 2.11 DNase I footprinting experiment .....	53
Figure 2.12 IC <sub>50</sub> of polyamides against HIV-1 integrase.....	54
Figure 2.13 Inhibition of strand transfer reactions.....	55

Table 2.1	Equilibrium association constants for M-MuLV LTR .....	29
Table 2.2	DNA melting with internal and terminal polyamide binding sites.....	43
Table 2.3	Equilibrium association constants for HIV-1 U5 LTR .....	53

### CHAPTER THREE

Figure 3.1	HER2 signaling pathway.....	65
Figure 3.2	Sequence of the HER2 promoter.....	67
Figure 3.3	Overview of polyamides targeted to the HER2 promoter .....	69
Figure 3.4	TATA Binding Protein .....	70
Figure 3.5	Polyamides targeted to the HER2 promoter in this study.....	72
Figure 3.6	DNase I footprinting experiments.....	75
Figure 3.7	Comparison of two different binding sites.....	77
Figure 3.8	TBP electrophoretic mobility shift assays .....	80
Table 3.1	Equilibrium association constants for HER2 promoter .....	74
Table 3.2	Equilibrium association constants for M-MuLV LTR .....	76

### CHAPTER FOUR

Figure 4.1	Structures of polyamide-Bodipy FL conjugates.....	93
Figure 4.2	Conjugate <b>1</b> in live SKBR-3 cells and after methanol addition .....	96
Figure 4.3	Conjugate <b>1</b> and the dead cell stain Sytox Orange in Sf9 cells.....	96
Figure 4.4	Conjugate <b>2</b> in live Sf-9 cells and after methanol addition.....	97
Figure 4.5	Conjugates <b>1</b> and <b>3</b> in live CEM cells.....	98
Figure 4.6	Synthesis of polyamide-Bodipy conjugates .....	106
Figure 4.7	Fluorescence of Bodipy conjugate <b>4</b> .....	107
Figure 4.8	DNase I footprinting experiments.....	108
Figure 4.9	Conjugate <b>3</b> in live Hi-5 cells and conjugate <b>1</b> in live LnCap cells .....	112
Figure 4.10	Synthesis of Bodipy-cyclic polyamide conjugate <b>11</b> .....	113
Figure 4.11	Tandem dimer and one-to-one binding Bodipy conjugates.....	114
Figure 4.12	Conjugates <b>13-14</b> in live NB4 and Sf9 cells .....	115



Figure 4.13	Structures of CHL-polyamide-Bodipy conjugates .....	117
Figure 4.14	CHL conjugates <b>18-19</b> in live NB4 and primary human T-cells.....	118
Figure 4.15	Conjugate <b>23</b> in live LnCap and PC3 cells; small molecule conjugates ....	120
Figure 4.16	Structures of carrier peptide conjugates.....	122
Figure 4.17	Carrier peptide conjugates <b>24,26,28</b> in live Sf9 and Kc cells .....	123
Figure 4.18	Structures of footprinting compounds.....	126
Figure 4.19	DNase I footprinting experiments.....	127
Figure 4.20	DNase I footprinting experiments.....	128
Figure 4.21	Structure of disulfide linked R9 conjugate.....	129
Figure 4.22	Structures of Bodipy-cationic polyamide conjugates .....	131
Figure 4.23	Conjugate <b>41</b> in live CEM and Kc cells.....	132
Figure 4.24	PD-4 conjugates <b>47-48</b> in live Kc and Sf9 cells.....	133
Table 4.1	Cellular localization of polyamide-Bodipy conjugates .....	95
Table 4.2	Equilibrium association constants for Bodipy conjugates .....	110
Table 4.3	Screening of standard hairpin-Bodipy conjugates.....	111
Table 4.4	Confocal microscopy of different polyamide motifs.....	114
Table 4.5	Screening of polyamide-chlorambucil-Bodipy conjugates .....	119
Table 4.6	Screening of polyamide-Bodipy-small molecule conjugates .....	121
Table 4.7	Screening of polyamide-Bodipy-carrier peptide conjugates .....	124
Table 4.8	Equilibrium association constants for peptide conjugates .....	128
Table 4.9	Screening of cationic polyamide-Bodipy conjugates.....	132

## CHAPTER FIVE

Figure 5.1	Schematic of polyamide binding.....	143
Figure 5.2	Structures of polyamides with different C-terminal tails.....	145
Figure 5.3	Polyamide synthesis on oxime resin .....	146
Figure 5.4	DNase I footprinting experiments.....	149
Figure 5.5	Model of aminoxy acid unit in a polyamide.....	159
Figure 5.6	Synthesis of Boc-pyrrole-aminoxy acid.....	161

Figure 5.7	Solid-phase synthesis of primary amides from aminooxy linker.....	162
Figure 5.8	Aminooxy linker route to C-terminal cyclic tertiary amides.....	163
Figure 5.9	Products from reaction with Dp or reductive conditions.....	164
Figure 5.10	DNase I footprinting experiments.....	165
Table 5.1	Equilibrium association constants for truncated tails.....	151
Table 5.2	Equilibrium association constants for tertiary amide tails .....	166

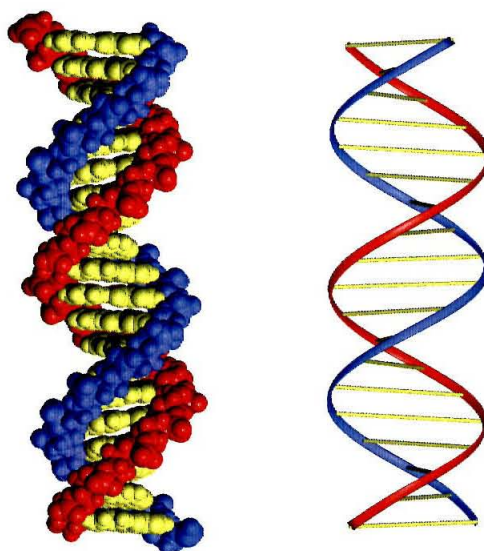
# **CHAPTER 1**

## **Introduction**

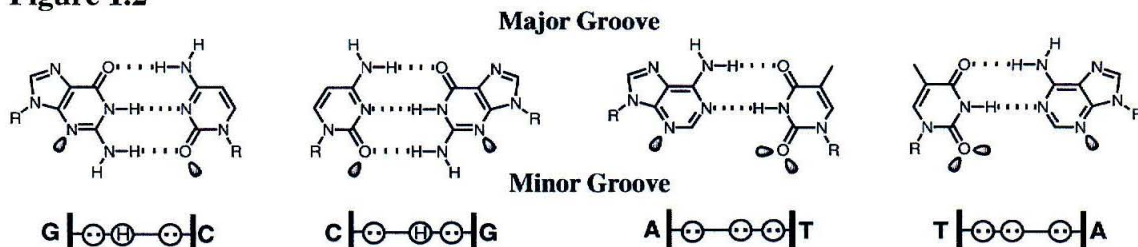
## Background

DNA is the storehouse of genetic information, containing all the information required for the growth and development of every organism.<sup>1</sup> DNA is composed of deoxyribose-phosphate polymers that display four heterocycles bases, adenine (A), thymidine (T), guanine (G), and cytosine (C). DNA is self-organized as a double-helical duplex of antiparallel strands, which are held together by Watson-Crick hydrogen bonding of A,T and G,C base pairs.<sup>2</sup> The recent sequencing of the human genome refers to deciphering the linear order of billions of these base pairs which code for 30,000-40,000 genes.<sup>3,4</sup> Understanding the complex interrelationships between these genes and their protein products is the central question in molecular biology and human medicine. Aberrant regulation of gene expression, through mutations or the intervention of pathogens, is responsible for a variety of disease states.<sup>5</sup> Synthetic ligands which can affect gene expression by binding to specific predetermined DNA sequences will be powerful tools for our growing understanding of the human genome, and potentially will be valuable therapeutic resources.<sup>6</sup>

**Figure 1.1**



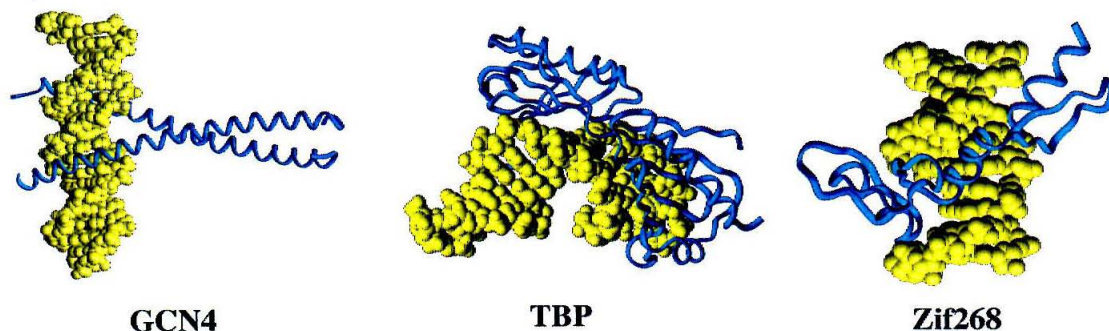
**Figure 1.1.** CPK model (left) and ribbon representation (right) of B-form duplex DNA. The sugar phosphate backbone is shown in either blue or red. The base pairs are in yellow.

**Figure 1.2**

**Figure 1.2.** The edges of the base pairs present different hydrogen bond donor and acceptor combinations to the major and minor grooves in the DNA double helix. The top of the bases, as drawn projects into the major groove and the bottom projects into the minor groove. Below the base pairs, the minor groove is shown in schematic representation, where circles with dots represent lone pairs of N(3) of purines (A,G) and O(2) of pyrimidines (C,T), and circles containing an H represent the 2-amino group of G.

## DNA Recognition

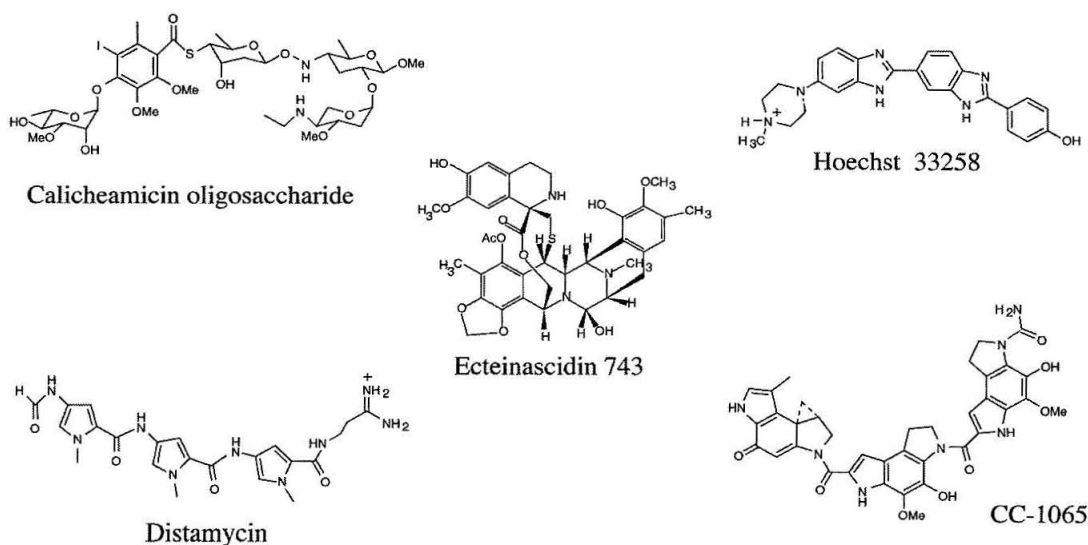
The common B-form of DNA is characterized by a wide (12 Å) and shallow major groove and a narrow (4-6 Å) and deep minor groove (Figure 1.1).<sup>7</sup> In addition, sequence-dependent structural variations, conformational properties, and solvent and counterion organization can distinguish local DNA structures.<sup>7</sup> Individual sequences are distinguished by the pattern of hydrogen bond donors and acceptors present at the edges of the base pairs (Figure 1.2). Proteins which recognize DNA take advantage of specific hydrogen bonding or van der Waals contacts with functional groups in the grooves, Coulombic attraction to the negatively charged phosphodiester backbone or to the

**Figure 1.3**

**Figure 1.3.** Representative X-ray crystal structures of DNA binding proteins.<sup>8-10</sup> Proteins contact the DNA via interactions with the bases in the major (GCN4, Zif268) and/or minor groove (TBP), as well as the sugar phosphate backbone. Some, such as TBP, cause a large distortion from common B-form DNA.

electrostatic potential in the grooves, and/or intercalation of aromatic functional groups between the DNA bases.<sup>8</sup> DNA binding proteins adopt a variety of structural motifs for sequence-specific recognition including zinc finger<sup>9</sup> and leucine zipper<sup>11</sup> motifs (Figure 1.3). Specificity for target sites is achieved through specific noncovalent interactions between the protein side chains and the nucleobases and phosphates of the DNA. However, no single motif exists that represents a general amino acid-base pair code for all DNA sequences.<sup>12</sup> Recognition by some proteins, such as the ubiquitous transcription factor TBP,<sup>10</sup> involves large distortions of the target DNA from its common B-form structure, including bending and unwinding of the helix.

**Figure 1.4**

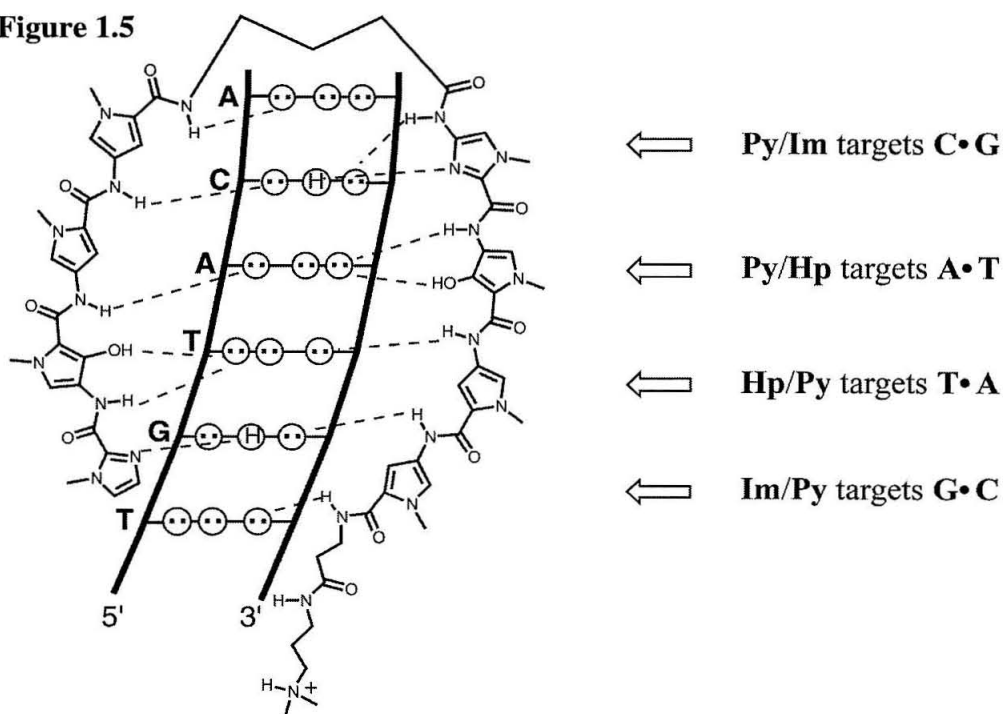


**Figure 1.4.** Structures of four natural products and a synthetic compound (Hoechst 33258) which bind in the minor groove of DNA. Note the similar crescent shape of distamycin, CC-1065 and Hoechst 33258.

Many small molecule natural products also recognize DNA, with varying degrees of sequence specificity by diverse binding modes including groove binding and intercalation.<sup>13-15</sup> The minor groove is a particular rich environment for recognition by small molecules (Figure 1.4).<sup>16,17</sup> Many of these compounds, such as the calicheamicin

oligosaccharide,<sup>18,19</sup> CC-1065,<sup>20,21</sup> and Hoeschst 33258,<sup>22,23</sup> possess modular structures that have allowed chemists to synthesize numerous derivatives in order to understand and modify their sequence specificity and binding affinity. Distamycin,<sup>24</sup> which binds five base pair A,T rich sequences, is particularly well suited to modification, owing to its relatively simple structure composed of *N*-methylpyrrole carboxamides.<sup>22,25-28</sup>

**Figure 1.5**



**Figure 1.5.** A schematic representation of the polyamide pairing rules.

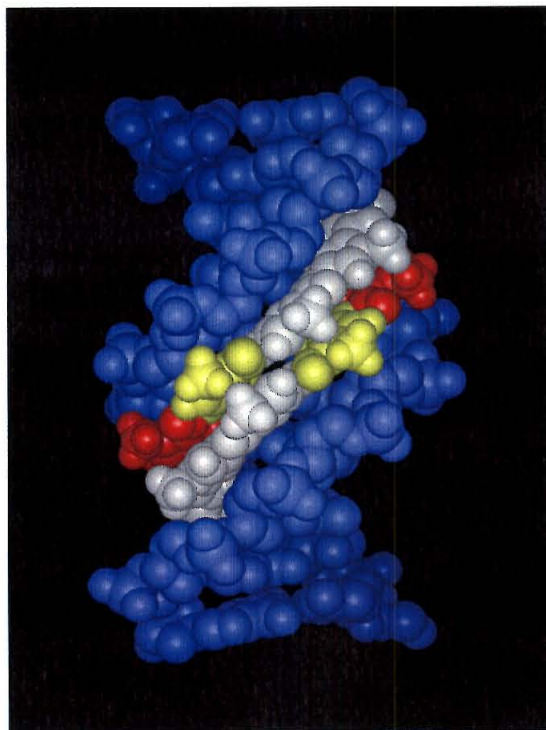
### DNA Recognition by Polyamides

A twenty year research effort in the laboratories of Prof. Peter B. Dervan have led to the development of a class of minor-groove binding small molecules which can recognize a large number of predetermined DNA sequences with affinities that rival natural transcription factors.<sup>6,29,30</sup> DNA binding polyamides contain *N*-methylpyrrole (Py) carboxamides, the structural repeat of distamycin, as well as *N*-methylimidazole (Im) and *N*-methyl-3-hydroxypyrrole (Hp) carboxamides. DNA recognition depends on side-by-



side amino acid pairings in the minor groove that stack the aromatic rings against each other and the walls of the groove allowing backbone amide hydrogens and the substituents at the 3-position of the Py, Im, and Hp residues to make specific contacts with the edges of the DNA base pairs. Each pair of polyamide residues is selective for a specific DNA base pair based on steric factors and the matching of hydrogen bond donor and acceptor functionalities (Figure 1.5). A pairing of Im opposite Py (Im/Py) targets a G•C base pair, while Py/Im targets C•G. A Py/Py pairing is degenerate, targeting both A•T and T•A base pairs. An Hp opposite a Py (Hp/Py) discriminates T•A from A•T, while Py/Hp targets A•T in preference to T•A and both of these from G•C and C•G. Footprinting, NMR, and X-ray structure studies validate these pairing rules for DNA minor groove recognition (Figure 1.6).<sup>29-31</sup>

**Figure 1.6**



**Figure 1.6.** Space filling representation of the polyamide dimer ImHpPyPy-β-Dp bound in the minor groove of DNA derived from a high-resolution X-ray co-crystal structure.<sup>31</sup>



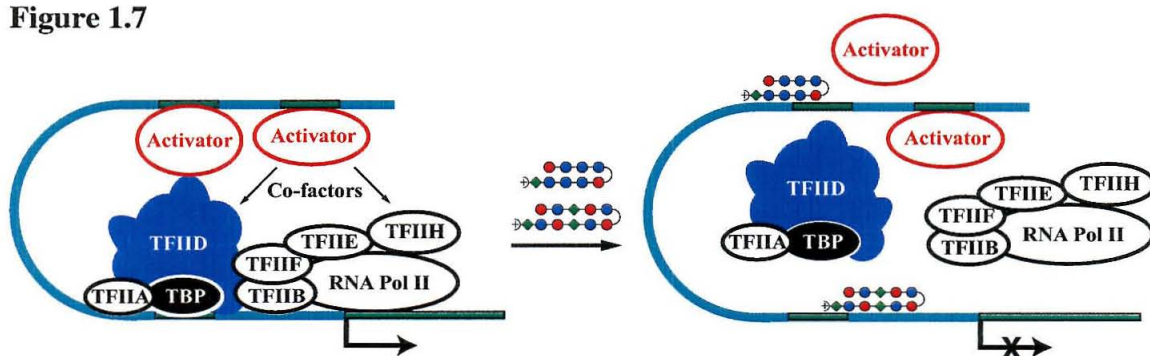
While many different recognition motifs based on pyrrole-imidazole polyamides are possible, the design known as the hairpin (Figure 1.5) has been widely used for biological applications.<sup>32</sup> The N- and C-terminal strands of Py/Im aromatic amino acids are connected by an alkyl amino acid--either  $\gamma$ -aminobutyric acid ( $\gamma$ ), or the chiral, amine-functionalized derivative (*R*)-2,4,-diaminobutyric acid ( $((R)^{H_2N}\gamma)$ --in an antiparallel orientation relative to each other and, generally, a 5'-3' N-C orientation relative to DNA. The linker between Py/Im strands is considered the "turn" of the hairpin, and alkyl units  $\beta$ -alanine ( $\beta$ ) and *N,N*-dimethylaminopropylamine (Dp) are typically used to form a C-terminal "tail." All of these alkyl units are specific for both A,T base pairs.  $\beta$ -alanine ( $\beta$ ) can also be used internally to form  $\beta/\beta$ ,  $\beta/Py$ , and  $\beta/Im$  pairs, which has extended the binding site size amenable to recognition by the hairpin polyamides and in certain cases results in increased affinity and specificity.<sup>33</sup> Tandem hairpin dimers have been produced which bind with high affinity and specificity.<sup>34,35</sup> The hairpin motif is also a versatile scaffold for the production of conjugates with diverse functions including sequence-specific DNA-alkylation<sup>36,37</sup> and recruitment of Topoisomerase I.<sup>38</sup>

## Gene Regulation

Hairpin polyamides have shown the ability to block transcription factors from binding to their DNA target sites and thereby inhibiting transcription (Figures 1.7 and 1.8).<sup>29,32</sup> In the first demonstration of this approach, an eight-ring hairpin polyamide was targeted to the binding site of the zinc finger transcription factor TFIIIA in the *Xenopus* 5S RNA gene promoter, at a sequence which coincides with a zinc finger-minor groove interaction (Figure 1.8).<sup>6</sup> Transcription by RNA polymerase III was suppressed in *in vitro* transcription assays and following treatment of whole *Xenopus* kidney-derived fibroblast cells with the match polyamide. Very recently it has been demonstrated that polyamides are able to inhibit the binding of certain zinc finger proteins (including Zif268, Figure 1.3) that bind without any minor groove contacts, presumably by an

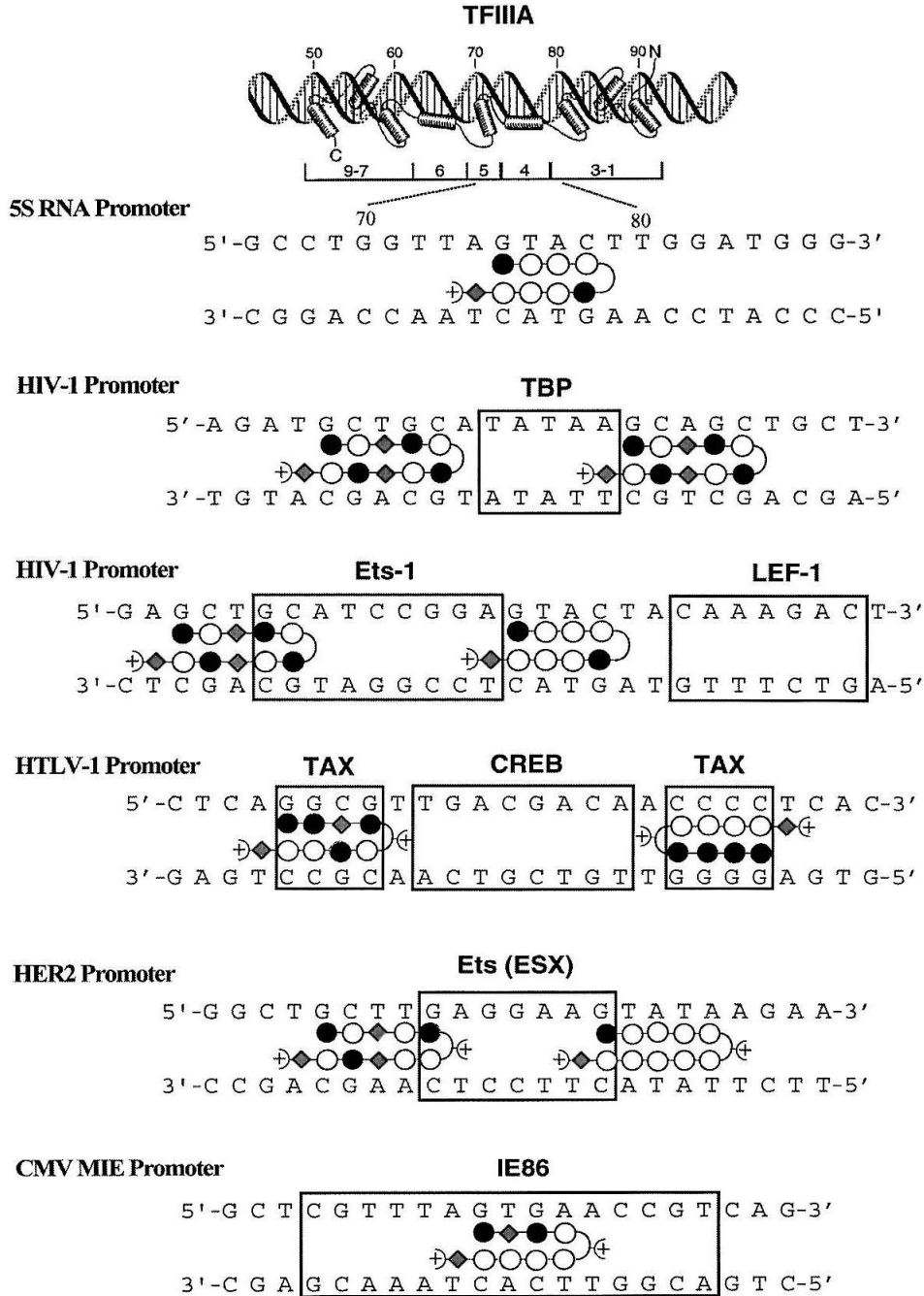
allosteric mechanism.<sup>39</sup> Allosteric interactions between the major and minor grooves were recently shown to be critical for inhibition of the major-groove binding protein UL9 by a minor-groove binding oligopyrrole small molecule (GLX, an indole bridged bis-netropsin).<sup>40</sup> However, minor-groove binding ligands including hairpin polyamides can also co-occupy DNA while certain proteins (such as GCN4, Figure 1.3) occupy the major groove.<sup>41</sup> GCN4 binding has been inhibited by hairpin polyamides containing “positive patches” targeting protein-phosphate contacts,<sup>42,43</sup> and by polyamide-intercalator conjugates.<sup>44</sup>

**Figure 1.7**



**Figure 1.7.** Inhibition of gene transcription by hairpin polyamides. A pair of hairpin polyamides targeted to the DNA sequences adjacent to the binding sites for Ets-1, LEF-1, and TBP inhibit assembly of the transcriptional machinery and transcription of the HIV-1 genome.<sup>45</sup> Polyamides are shown in ball-and-stick format. Red and blue circles represent imidazole and pyrrole residues, respectively.  $\beta$ -Alanine and the  $\gamma$ -turn are depicted as green diamonds and curved lines, respectively. The plus sign represents the dimethylaminopropylamide tail.

Minor-groove binding transcription factors, such as TBP (Figure 1.3), LEF-1, and TAX, have been excellent targets for polyamide inhibition, resulting in inhibition of RNA polymerase II transcription of targeted genes *in vitro* such as HIV-1 and HTLV-1 (Figure 1.8).<sup>45-47</sup> Interestingly, in the absence of TAX the polyamides shown on the HTLV-1 promoter or another polyamide targeted to the center of the CREB site, do not inhibit the association of major-groove binding CREB with DNA.<sup>47</sup> However, in the presence of TAX, which forms a complex with CREB, the polyamides shown inhibited CREB

**Figure 1.8**

**Figure 1.8.** Examples of gene promoters where transcription factor binding sites that have been targeted by hairpin polyamides, resulting in blocking transcription factor binding and inhibition of transcription by RNA polymerase III (5S RNA) and RNA polymerase II, or activation of transcription by blocking the repressor IE86.<sup>6,45-48</sup> Polyamides are shown in ball-and-stick format as in Figure 1.7, with Im as a filled circle, Py as an open circle, and  $\beta$ -alanine as a grey diamond.

binding. A similar result was obtained on the HIV-1 promoter with a polyamide targeting Ets-1 (Figure 1.8, polyamide on the left), which was able to inhibit binding of NF- $\kappa$ B to an adjacent DNA site by blocking the formation of a cooperative Ets-1•NF- $\kappa$ B•DNA ternary complex.<sup>49</sup> These results show that polyamides are able to inhibit proteins that they do not directly contact, and may provide a general method for inhibition of the association of major-groove binding transcription factors with their target DNA.

Proteins that have a majority of major-groove contacts and few or even one minor-groove contact have been efficiently inhibited by hairpin polyamides. The binding of two Ets transcription factors, Ets-1 and ESX, to the HIV-1 and HER2 promoters, respectively, has been efficiently inhibited by hairpin polyamides (Figure 1.8).<sup>45,46,49</sup> Ets proteins have a winged-helix-turn-helix motif and bind primarily in the major groove with additional phosphate contacts across the adjacent minor groove(s). Polyamide inhibition of the binding of NF- $\kappa$ B, which has one such contact across the minor groove, is likely due, at least in part, to allosteric interactions between the grooves.<sup>50</sup> However, in another case of polyamides inhibiting transcription factor binding with only one phosphate contact outside the major groove, synthetic modification of the bHLH domain of the *Drosophila* Deadpan showed that the inhibition depended on the specific lysine residue responsible for the phosphate contact.<sup>51</sup> Inhibition by two polyamides with distinct binding sites was then used to demonstrate unusual asymmetric binding of this homodimeric bHLH protein. This highlights the use of polyamide not only for inhibition but also for the study of transcription factor-DNA complexes.

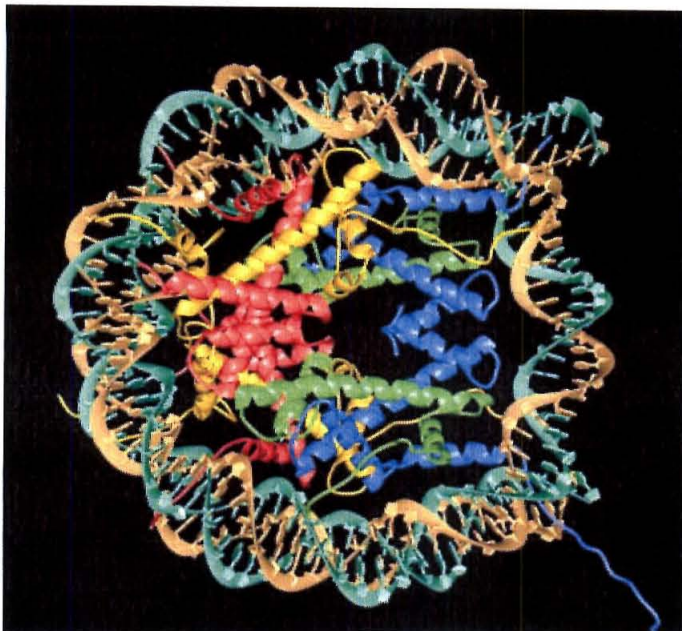
Recently, *in vitro* transcription inhibition has been demonstrated for a hairpin polyamide binding to sites that do not involve sequence-specific protein-DNA interactions. A “promoter scanning” experiment<sup>52</sup> was devised to probe the portion of the HIV-1 promoter around the TATA box, using a series of DNA constructs in which a hairpin polyamide binding site was placed at various distances from the TATA box.<sup>53</sup> As expected, for the constructs with polyamide sites proximal to the TATA box, TBP

binding and transcription by RNA polymerase II were inhibited. However, repression of transcription was also observed for sites removed from the TATA box, which did not show inhibition of isolated TBP binding. As seen in Figure 1.7, the RNA polymerase II transcriptional machinery is a huge complex, which occupies at least 40 base pairs of DNA up stream of the transcription start site (the TATA box is at -28 of the HIV-1 promoter). Inhibition was traced to polyamides blocking the TFIID•TFIIA•DNA ternary complex, which is not affected by the DNA sequence, based on prior mutagenesis results. This study suggests that the range of potential target sites for transcription inhibition by polyamides may be considerably larger than was previously thought.

In addition to inhibition, polyamides have also been shown to activate transcription. This was accomplished with standard hairpin polyamides by inhibiting the binding of a transcriptional repressor protein, IE86, to its cognate site in the viral CMV major intermediate early promoter (Figure 1.8).<sup>48</sup> *In vitro* transcriptional activation has also been accomplished by hairpin polyamides linked to known activation peptides.<sup>54,55</sup> A peptidic dimerization domain in the original design has been replaced by successively shorter linkers, down to eight atoms, without a significant loss of activation. Promoter-specific activation over 30-fold above basal levels has been observed.<sup>55</sup> Since polyamides can target large variety of sequences with high affinity, this approach has great potential for gene-specific activation by small molecules.

In order to activate or inhibit transcription *in vivo*, polyamides will have to interact with DNA that is not free in solution but organized in the cell nucleus in chromatin. The fundamental repeat of chromatin is the nucleosome, which consists the nucleosome core particle (NCP) and 20-80 base pairs of linker DNA. The NCP consists of two tight superhelical turns of DNA (147 base pairs) wrapped around a disc-shaped core composed of eight histones (Figure 1.9).<sup>56</sup> The histones are not simply structural proteins, they are intimately involved in virtually all cellular processes involving DNA, through direct interactions with other proteins and by affecting the accessibility of



**Figure 1.9**

**Figure 1.9.** Representation of the X-ray crystal structure of the NCP used for studying the binding of hairpin polyamides to nucleosomal DNA.<sup>56</sup> Due to their flexibility, only one of the histone tails showed up in the crystal structure.

specific sequences. In particular, reversible acetylation of the histone tails plays a major role in transcription.<sup>57</sup> Surprisingly, most of the DNA on an isolated NCP was found to be accessible to polyamide binding, including sites where the minor groove faces into the histone octamer, although affinity was reduced at those sites.<sup>58</sup> The only sites which were found to be inaccessible to polyamide recognition were blocked by the passage of the histone tails through the minor groove. The nucleosomes remained fully folded upon polyamide binding and the hairpins did not interfere with reconstitution of the NCPs from free DNA and histones. Recently similar results on the accessibility of minor groove sites both facing away from and into the histone core without disruption of the NCP were obtained for the smaller minor-groove binder Hoechst 33258 (Figure 1.4).<sup>59</sup> These studies suggest that the vast majority of DNA sites in nucleosomes will be available for recognition by hairpin polyamides. However, it should be noted that nucleosomes are only the first level of chromatin organization.

Polyamides have been shown to regulate gene expression *in vivo*. The two most powerful demonstrations are the inhibition of HIV-1 replication in human cells by polyamides targeted to the TBP, Ets-1, and Lef-1 sites on the HIV-1 promoter by hairpin polyamides (Figures 1.7 and 1.8),<sup>45</sup> and the activation and repression of selected genes in living *Drosophila melanogaster* by one-to-one binding polyamides targeted to highly repeated satellite DNA sequences.<sup>60,61</sup> In the latter study Laemmli and co-workers demonstrated that monomeric and dimeric one-to-one binding polyamides showed specificity for satellite repeat sequences in DNase I footprinting assays, as well as in isolated chromosomes and nuclei as measured by epifluorescence microscopy visualization of fluorescently tagged polyamides. They also showed that monomeric compounds mediated chromatin opening of heterochromatic satellite DNA containing the target repeat.<sup>60</sup> Next, the monomeric polyamides were fed to developing flies, which proceeded to show specific gain- and loss-of-function phenotypes based on the chromatin opening of the targeted DNA satellites.<sup>61</sup> Laemmli and co-workers have followed up on these studies by developing fluorescently tagged tandem hairpin dimers targeted to insect and vertebrate telomere repeat sequences.<sup>62</sup> These were shown to be excellent probes for staining telomeres and rapidly estimating telomere length.

In the HIV-1 study, the laboratories of Dervan, Gottesfeld, and Mosier showed that a combination of two hairpin polyamides which inhibited the binding of three transcription factors and *in vitro* transcription of the HIV-1 promoter, reduced HIV-1 replication 99.9% (below the detection limit, similar to positive control AZT) in isolated human peripheral blood lymphocytes infected with a T-cell tropic strain of HIV-1.<sup>45</sup> Individually the polyamides were also quite effective, causing 60% and 80% reduction in virus. Mismatch polyamides showed no inhibition, and the mRNA levels of seven other genes with different proximal sequences to their TATA boxes were not effected by the polyamide treatment. The concentration of polyamide required for maximal inhibition, 1  $\mu$ M of each, was not appreciably toxic to the cells.

## Scope of this Work

This thesis describes work examining DNA binding polyamides in biological systems. Chapters 2 and 3 are both conceptually related to the HIV-1 study with hairpin polyamides. Chapter 2 describes the development of polyamides as integrase inhibitors, which represents a complimentary attack on the retroviral life-cycle. Potent inhibitors of M-MuLV (a retrovirus related to HIV-1) integration in cell-free assays are discussed in Chapter 2A. Chapter 2B presents a series of hairpin polyamides designed to inhibit HIV-1 integration. Following on the inhibition of a viral gene in human cells, the Dervan laboratory sought to inhibit the transcription of an endogenous human gene in human cells. HER2, a gene overexpressed in human breast cancer, was chosen as a target. Chapter 3 describes a series of polyamides targeting the TATA box of the HER2 promoter. Chapter 4 describes confocal microscopy studies of polyamide-fluorescent dye conjugates which followed from failure to achieve HER2 inhibition *in vivo*. Chapter 4A suggests that hairpin polyamides are indeed cell-permeable, but that nuclear localization does not occur in many cell lines. Efforts toward modified polyamides with enhanced nuclear localization properties are described in Chapter 4B. Chapter 5 describes efforts to increase the number of sequences amenable to high-affinity recognition by hairpin polyamides by alteration of the C-terminal tail. Chapter 5A presents the development of conditions for polyamide solid-phase synthesis on a new resin, which allows for the generation of “truncated tail” polyamides. Chapter 5B describes the original route to one of these compounds, during the development of which an unexpected reaction was uncovered that led to entirely different C-terminal tails.



## References

- (1) Lewin, B. *Genes V*; Oxford University Press: New York, 1994.
- (2) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, M.; Fratini, A. V.; Kopka, M. L. *Science* **1982**, *216*, 475.
- (3) Consortium, I. H. G. S. *Nature* **2001**, *409*, 860.
- (4) Venter, J. C. et al. *Science* **2001**, *291*, 1304.
- (5) Tjian, R. *Sci. Am.* **1995**, *2*, 54.
- (6) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205.
- (7) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (8) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053-1095.
- (9) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809.
- (10) Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512-520.
- (11) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
- (12) Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579.
- (13) Gao, X. L.; Mirau, P.; Patel, D. J. *J. Mol. Biol.* **1992**, *223*, 259-279.
- (14) Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, *225*, 445-456.
- (15) Guan, Y.; Sakai, R.; Rinehart, K. L.; Wang, A. H.-L. *J. Biomol. Struct. Dyn.* **1993**, *10*, 793-818.
- (16) Wemmer, D. E.; Dervan, P. B. *Current Opinion in Structural Biology* **1997**, *7*, 355-361.
- (17) Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 463-493.

- (18) Liu, C.; Smith, B. M.; Ajito, K.; Komatsu, H.; GomezPaloma, L.; Li, T. H.; Theodorakis, E. A.; Nicolaou, K. C.; Vogt, P. K. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 940-944.
- (19) Kalben, A.; Pal, S.; Andreotti, A. H.; Walker, S.; Gange, D.; Biswas, K.; Kahne, D. *J. Am. Chem. Soc.* **2000**, *122*, 8403-8412.
- (20) Boger, D. L.; Schmitt, H. W.; B.E., F.; Hendrick, M. P. *J. Org. Chem.* **2001**, *66*, 6654-6661.
- (21) Boger, D. L.; Stauffer, F.; Hendrick, M. P. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2021-2024.
- (22) Satz, A. L.; Bruice, T. C. *Acc. Chem. Res.* **2002**, *35*, 86-95.
- (23) Minehan, T. G.; Gottwald, K.; Dervan, P. B. *Helv. Chim. Acta.* **2000**, *83*, 2197-2213.
- (24) Arcamone, F.; Penco, S.; Prezzi, P. G.; Nicoletta, V.; Pirelli, A. *Nature* **1964**, *203*, 1064.
- (25) Bailly, C.; Chaires, J. B. *Bioconjugate Chem.* **1998**, *9*, 513-538.
- (26) Boger, D. L.; Dechantsreiter, M. A.; Ishii, T.; Fink, B. E.; Hendrick, M. P. *Bioorg. Med. Chem.* **2000**, *8*, 2049-2057.
- (27) O'Hare, C. C.; Mack, D.; Tandon, M.; Sharma, S. K.; Lown, J. W.; Kopka, M. L.; Dickerson, R. E.; Hartley, J. A. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 72-77.
- (28) Dyatkina, N. B.; Roberts, C. D.; Keischer, J. D.; Dai, Y.; Nadherny, J. P.; Zhang, W.; Schmitz, U.; Kongpachith, A.; Fung, K.; Novikov, A. A.; Lou, L.; Velligan, M.; Khorlin, A. A.; Chen, M. S. *J. Med. Chem.* **2002**, *45*, 805-817.
- (29) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.
- (30) Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688-693.
- (31) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111-115.
- (32) Gottesfeld, J. M.; Turner, J. M.; Dervan, P. B. *Gene Expression* **2000**, *9*, 77-91.

- (33) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.
- (34) Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem.-Eur. J.* **1999**, *5*, 975-983.
- (35) Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218-3228.
- (36) Chang, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 4856-4864.
- (37) Wurtz, N. R.; Dervan, P. B. *Chem. Biol.* **2000**, *7*, 153-161.
- (38) Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2001**, *123*, 8657-8661.
- (39) Nguyen, D.H.; Dervan, P.B.; Pabo, C.O. Unpublished Results.
- (40) Kwok, Y.; Zhang, W.; Schroth, G. P.; Liang, C. H.; Alexi, N.; Bruice, T. W. *Biochemistry* **2001**, *40*, 12628-12638.
- (41) Oakley, M. G.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10969-10975.
- (42) Bremer, R. E.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1998**, *5*, 119-133.
- (43) Bremer, R. E.; Wurtz, N. R.; Szewczyk, J. W.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2093-2103.
- (44) Fechter, E.; Dervan, P.B. Unpublished results.
- (45) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895.
- (46) Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A. *J. Biol. Chem.* **2000**, *275*, 24246-24254.
- (47) Lenzmeier, B. A.; Baird, E. E.; Dervan, P. B.; Nyborg, J. K. *J. Mol. Biol.* **1999**, *291*, 731-744.
- (48) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Ghazal, P.; Dervan, P. B.; Gottesfeld, J. M. *Biochemistry* **1999**, *39*, 10801-10807.
- (49) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Graves, B. J.; Gottesfeld, J. M. *J. Biol. Chem.* **1999**, *274*, 12765-12773.
- (50) Wurtz, N.R.; Dervan, P.B.; Baltimore, D. Unpublished Results.

- (51) Winston, R. L.; Ehley, J. A.; Baird, E. E.; Dervan, P. B.; Gottesfeld, J. M. *Biochemistry* **2000**, *39*, 9092-9098.
- (52) McBryant, S. J.; Baird, E. E.; Trauger, J. W.; Dervan, P. B.; Gottesfeld, J. M. *J. Mol. Biol.* **1999**, *286*, 973-981.
- (53) Ehley, J. A.; Melander, C.; Herman, D.; Baird, E. E.; Ferguson, H. A.; Goodrich, J. A.; Dervan, P. B.; Gottesfeld, J. M. *Mol. Cell. Biol.* **2002**, *22*, 1723-1733.
- (54) Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3930.
- (55) Ansari, A. Z.; Mapp, A. K.; Nguyen, D. H.; Dervan, P. B.; Ptashne, M. *Chem. Biol.* **2001**, *2001*, 583-592.
- (56) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251-260.
- (57) Narliker, G. J.; Fan, H.-Y.; Kingston, R. E. *Cell* **2002**, *108*, 475-487.
- (58) Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. *J. Mol. Biol.* **2001**, *309*, 615-629.
- (59) Leslie, K. D.; Fox, K. R. *Biochemistry* **2002**, *41*, 3484-3497.
- (60) Janssen, S.; Durussel, T.; Laemmli, U. K. *Mol. Cell* **2000**, *6*, 999-1011.
- (61) Janssen, S.; Cuvier, O.; Muller, M.; Laemmli, U. K. *Mol. Cell* **2000**, *6*, 1013-1024.
- (62) Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218-3228.

## **CHAPTER 2**

### **Inhibition of Retroviral Integration by DNA Binding Polyamides**

*The text of this chapter was taken in part from a manuscript submitted for publication coauthored with Professor Peter Dervan (Caltech) and Fan Yang, Rodrigo Villanueva, and Professor Monica Roth (Robert Wood Johnson Medical School).*

(Yang, F.; Belitsky, J.M.; Villanueva, R.M.; Dervan, P.B.; Roth, M.J. "Inhibition of Moloney murine leukemia virus integration using polyamides targeting the long-terminal repeat (LTR) sequences" *Biochemistry*, submitted 2002.)

## Abstract

The retroviral integrase (IN), which carries out the integration of the viral DNA into the host genome, is a potential target for antiviral therapy. Both IN and the DNA sequences at the viral long terminal repeat (LTR) are required for the integration function. Chapter 2A describes the investigation of a series of minor groove binding hairpin polyamides targeting sequences within terminal inverted repeats of the Moloney murine leukemia virus (M-MuLV) LTR which were synthesized as potential inhibitors of integration. Using cell-free *in vitro* integration assays, polyamides targeting the conserved CA dinucleotide with cognate sites closest to the terminal base pairs were effective at blocking 3' processing but not strand transfer. Polyamides which efficiently inhibited 3' processing and strand transfer targeted the LTR sequences through position 9. Polyamides that inhibited integration were effective at nanomolar concentrations and showed subnanomolar affinity for their cognate LTR sites. These studies develop the use of polyamides for antiretroviral therapy.

Chapter 2B describes an extension of these studies to the inhibition of HIV-1 integration. Hairpin polyamides targeting sequences within the U5 LTR of HIV-1 were synthesized and shown to bind their cognate sites with subnanomolar affinity. These compounds were found to be submicromolar inhibitors of 3' processing and strand transfer in preliminary cell-free integration assays. Comparison with the M-MuLV results and potential approaches to integrase inhibitors with greater potency are discussed.

## Chapter 2A:

### Inhibition of Moloney murine leukemia virus integration using polyamides targeting the long-terminal repeat (LTR) sequences

#### Introduction

Emergence of drug-resistant strains of human immunodeficiency virus (HIV) calls for the identification of new antiviral targets. The retroviral IN is a rational target for antiviral therapy both because it is essential for the viral replication and because it has no cellular counterpart. The integrase catalyzes the integration of viral DNA into the host genome. In addition to the integrase, the DNA sequences present in the long terminal repeat (LTR) at the ends of the linear viral DNA are also required for integration.<sup>1,2</sup> The integration process can be divided into two steps. First, IN cleaves two nucleotides from the 3' ends of both LTR termini, exposing the conserved a 5'-CA-3' dinucleotide. Second, the 3' processed-ends of the viral DNA are joined to the host DNA in a coordinated fashion. *In vitro* assays have been developed to recapitulate both steps of integration by using purified integrase protein and short oligonucleotide duplexes mimicking the viral LTR.<sup>3-5</sup> In addition, purified integrase is able to carry out an *in vitro* concerted two-end integration reaction using the LTR oligonucleotide as donor and plasmid DNA as target.<sup>6,7</sup>

Multiple potential IN inhibitors have recently been identified and classified into four groups based on the proposed sites of action. Inhibitors that target the metal binding site include salicylhydrazides<sup>8</sup> and styrylquinolines.<sup>9</sup> The proposed mode of action of these compounds is through chelating the divalent metal ions required by IN. A second class targets a potential nucleotide binding site and includes 5-N<sub>3</sub>-AZTMP,<sup>10</sup> pIsopdApdC,<sup>11</sup> and guanosine quartets.<sup>12-15</sup> These inhibitors are proposed to compete with the viral DNA binding site. Alternative nucleic acid binding sites act as potential targets. Two inhibitors in this class, 5-CITEP<sup>16</sup> and naphthalene disulfonate,<sup>17</sup> were

found to bind to the active site of the integrase by structural studies. Compounds containing diketo acids are recently reported to specifically block strand transfer by competing for the target binding site.<sup>18,19</sup>

The viral DNA is also a target for inhibitors. Inhibitors in this class include DNA intercalators<sup>20</sup> as well as DNA minor-groove binders such as distamycin dimers.<sup>21-24</sup> Of particular relevance to this study, the minor groove binder netropsin has been shown to inhibit Moloney murine leukemia virus (M-MuLV) IN at micromolar concentrations.<sup>25</sup> Netropsin binds to A + T rich sequences of 4-7 base pairs in length. The terminal seven nucleotides of the M-MuLV LTR (5'-AATGAAA) are A + T rich, and contain sequences especially favorable for netropsin binding.<sup>25</sup>

Studies with modified LTRs containing nucleotide analogues indicate that the IN does contact the LTR minor groove.<sup>26</sup> However, the molecular details of integration remain elusive. Integration is a complex and dynamic process, which despite intense efforts,<sup>27-29</sup> has resisted direct structural characterization. There is evidence that different IN-LTR interactions are important for the individual steps of integration.<sup>18,30</sup> In the absence of a crystal structure, DNA binding molecules may shed light on the molecular interactions of integrase with the viral LTR. In this regard, it would be especially useful to place minor groove DNA binders at precise but incrementally different positions within the 13 base pair IN recognition sequence, 5'-AATGAAAGACCCC-3'.

Hairpin polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamides are small molecules that bind DNA in the minor groove with high affinity in a sequence-specific manner.<sup>31,32</sup> Based on pairing rules, polyamides can target a large number of predetermined DNA sequences. This level of target flexibility and precision, which allows unambiguous recognition of DNA sites with single base pair resolution, is not available to classical DNA binding agents. Polyamides have been shown to block transcription factors from binding to their cognate recognition sequences, thus inhibiting transcription.<sup>33-35</sup> Polyamides targeted to TBP and Lef-1 binding sites in the HIV-1



promoter have been shown to inhibit viral replication in human cells.<sup>33</sup> If polyamides can effectively target the LTR termini, they should act as integrase inhibitors in analogy to netropsin, which would represent a complimentary attack on the retroviral life cycle. Furthermore, a series of polyamides that recognize distinct LTR sites with high affinity would be molecular probes for understanding the mechanism of integration.

In this study, the ability of a series of polyamides (**1-8**), including two mismatch controls, to inhibit the *in vitro* integration activities of M-MuLV IN was determined. Inhibitors effective at nanomolar concentrations fell into two classes, based on the position of their cognate sites within the M-MuLV LTR. The characterization of these first generation polyamide inhibitors is presented.

## Experimental Procedures

**Materials.** Crude ( $\gamma$ -<sup>32</sup>P) ATP (7,000 Ci/mmol) was purchased from ICN. T4 polynucleotide kinase, T4 DNA ligase, DNase I and restriction enzymes were obtained from New England Biolabs. Ni<sup>2+</sup>-nitrilotriacetic acid agarose was purchased from Qiagen. Target plasmid pGEM-3Zf(+) was purchased from Promega.

**Polyamides.** All polyamides were synthesized by solid-phase methods as previously described,<sup>36-38</sup> and characterized by analytical HPLC and MALDI-TOF mass spectrometry. ImPyPyPyIm- $\gamma$ -PyPyPyPyPy- $\beta$ -Dp (**1**) [M+H] 1466.99, 1466.57 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)-<sup>H2N</sup> $\gamma$ -PyPyPyPyPy- $\beta$ -Dp (**2**) [M+H] 1430.94, 1430.67 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)-<sup>H2N</sup> $\gamma$ -PyPy- $\beta$ -PyPy- $\beta$ -PyPy- $\beta$ -Dp (**3**) [M+H] 1695.08, 1694.79 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)-<sup>H2N</sup> $\gamma$ -PyImPyPyPy- $\beta$ -Dp (**4**) [M+H] 1431.48, 1431.66 calculated for [M+H]. BzImPyPy-(R)-<sup>H2N</sup> $\gamma$ -PyPyPyPy- $\beta$ -Dp (**5**) [M+H] 1233.6, 1233.57 calculated for [M+H]. BzPyPyPy-(R)-<sup>H2N</sup> $\gamma$ -PyImPyPy- $\beta$ -Dp (**6**) [M+H] 1233.54, 1233.57 calculated for [M+H]. BzPyPyPy-(R)-<sup>H2N</sup> $\gamma$ -

PyPyImPy- $\beta$ -Dp (7) [M+H] 1233.6, 1233.57 calculated for [M+H]. ImPyPyIm-(R)- $^{H^{2N}}\gamma$ -PyPyImPy- $\beta$ -Dp (8) [M+H] 1238.73, 1238.58 calculated for [M+H].

**DNase I footprinting.** DNase I footprinting reactions were performed as previously described,<sup>39</sup> using a 183 base pair 5'- $^{32}\text{P}$ -labeled DNA fragment containing one copy of the M-MuLV LTR terminus. The labeled fragment was generated by the PCR method<sup>34,40</sup> using primers 5'-AGACAGGATATCAGTGGTCCA and 5'-CATGCCTTGCAAAATGGCGTT, corresponding to positions 10968-10988 and 43-63, respectively, of the M-MuLV linear provirus-containing plasmid, NCA-C,<sup>41</sup> which was used as the template. Primers were prepared and purified by the Caltech Biopolymer Synthesis Facility. The sequence of the labeled fragment is listed here with the terminal 13 base pairs of the M-MuLV LTR shown in bold: 5'- $^{32}\text{P}$  CATGCCTTGCAAAATGGCGTTACTTAAGCTAGCTAGCTTGCCAAACCTACAGGT**GGGGTCTTTCATTCC**CCCCTTTTCTGGAGACTAAATAAAATCTTTTATTTTATCTATGGCTCGTACTATAGGCTTCAGCTGGTGATATTGTTGAGTCAAAATAGAGCCTGGACCACTGATATCCTGTCT.

**Oligonucleotides.** DNA oligonucleotides used as substrates were prepared by the University of Medicine and Dentistry of New Jersey Biochemistry Department DNA Synthesis Facility and purified by electrophoresis on 20% denaturing polyacrylamide gels. Oligonucleotides used in this study are referred to by their synthesis numbers and were labeled with ( $\gamma$ - $^{32}\text{P}$ ) ATP by kinase reaction as previously described.<sup>42</sup> Oligonucleotides 2783 (5'-GTCAGCGGGGGTCTTTCATT) and its complementary strand 2785 (5'-AATGAAAGACCCCGCTGAC) were used for 3' processing assay. Oligonucleotides 2784 (5'-GTCAGCGGGGGTCTTTCAT) and its complementary strand 2785 were used for strand transfer and concerted two-end integration assays.

**Purification of M-MuLV integrase.** Recombinant M-MuLV integrase (WT IN) containing a hexahistidine tag were expressed in *Escherichia coli* BL21(DE3) (Novagen) and purified by  $\text{Ni}^{2+}$ -nitrilotriacetate agarose chromatography (Qiagen) as previously described.<sup>42</sup>

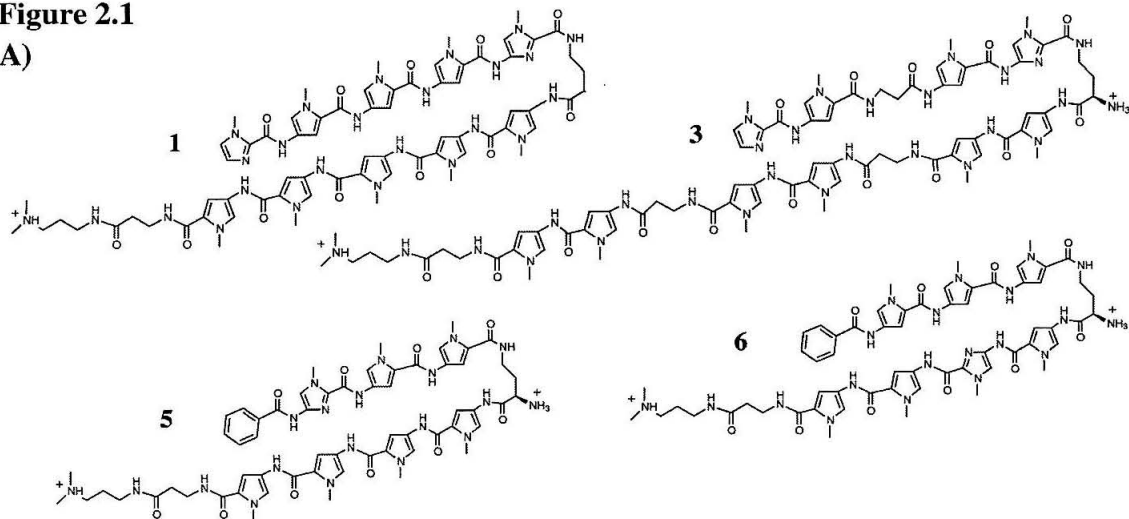
**In vitro assays.** Strand transfer and 3' processing reactions were performed as previously described.<sup>42</sup> The reaction buffer contains 20 mM morpholine-ethanesulfonic acid (MES, pH 6.2), 10 mM DTT, 10 mM  $\text{MnCl}_2$ , 10 mM KCl and 10% glycerol. The condition for integration into an exogenous target was the same as that of the strand transfer reactions except that 200 mM KCl and 10% DMSO (dimethylsulfoxide) were added. The LTR oligonucleotide (2783 or 2784) was 5' labeled by T4 polynucleotide kinase and mixed with complementary strand at a ratio of 1:2. The oligonucleotides were annealed by heating for three minutes at 95° C and then cooling to room temperature. Typically, one reaction mixture (30  $\mu\text{l}$ ) contains 1 pmol of labeled LTR, 0.6  $\mu\text{g}$  of target plasmid DNA and 20 pmol of IN protein. After the pre-incubation of the LTR and the polyamides, the IN protein was added and incubated on ice for 5 minutes and 37° C for 5 minutes. The target DNA and KCl were then added. The reactions were incubated at 37° C for two hours and stopped by addition of 10 mM EDTA, pH 8.0, 0.5% SDS and 100  $\mu\text{g}/\text{ml}$  proteinase K and incubated at 37°C for one hour. 10  $\mu\text{l}$  of the reaction was subject to electrophoresis on a 1% agarose gel. After gel electrophoresis, the gel was dried and exposed to Kodak X-Omat Blue XB-1 film. In all reactions, the polyamides were mixed with the LTR substrate first and incubated at room temperature for 18 hours before adding the IN protein, unless indicated otherwise. Films were scanned and quantified with a densitometric plot using the Scion Image 1.62c program (based on NIH image).

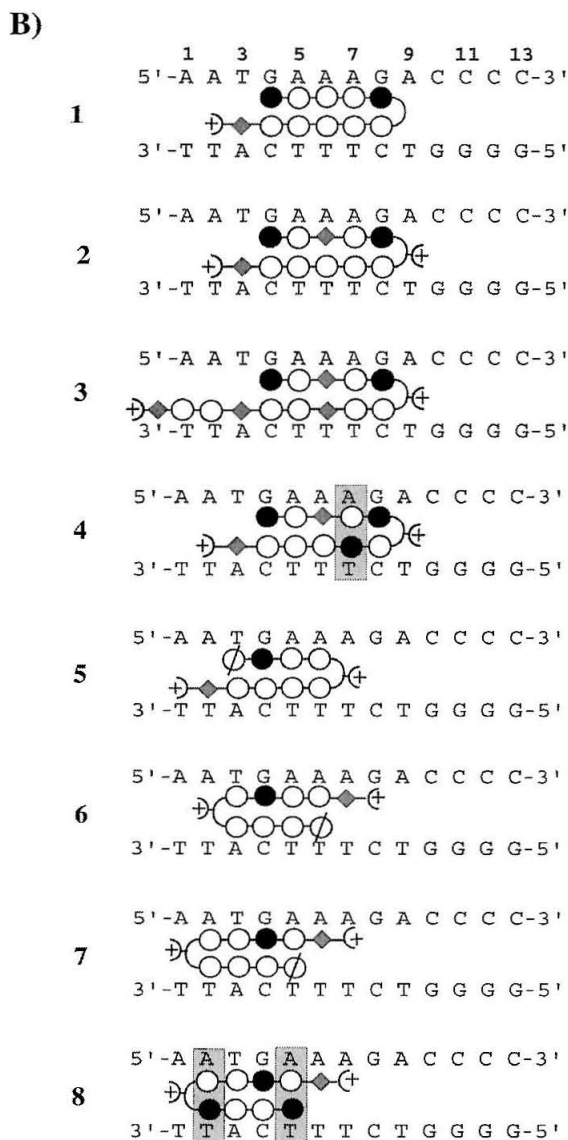
## Results

**Polyamide design.** The termini of the MuLV LTRs including the conserved CA::GT dinucleotide, consist of a 13 base pair (bp) inverted repeat sequence (5'-AATGAAGACCCC). The MuLV U5 and U3 termini sequences are therefore identical. This degree of conservation of the LTR termini is not found in HIV-1 and Rous sarcoma virus (RSV) genomes. This facilitates the generation of sequence directed polyamide inhibitors which would simultaneously recognize both ends of the viral DNA. A series of polyamides which target the MuLV LTR were generated (Figure 2.1). The pairing rules are outlined below for hairpin polyamide binding to DNA.<sup>31</sup> A pyrrole opposite an imidazole (Py/Im pairing) targets a C•G bp whereas an Im/Py pair binds a G•C bp. A Py/Py pair recognizes both A•T and T•A pairs. Internal  $\beta$ -alanines ( $\beta$ ) may be introduced to improve both flexibility and specificity of the polyamides.  $\beta/\beta$ ,  $\beta/\text{Py}$ , and  $\text{Py}/\beta$  pairs recognize both A•T and T•A pairs. The “turn” and “tail” residues,  $\gamma$ -aminobutyric acid ( $\gamma$ ), (*R*)-2,4,-diaminobutyric acid ((*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ ), unpaired  $\beta$ -alanine, and dimethylamino-propylamine (Dp), are also A•T and T•A specific. Recently a benzoyl group (Bz) has been shown to be a high affinity A•T and T•A specific N-terminal residue when paired with Py.<sup>38</sup>

**Figure 2.1**

**A)**





**Figure 2.1.** (A) Chemical structures of selected polyamides. (B) Ball-and-stick representations of the polyamides and their binding sites within the M-MuLV LTR. Numbering of the LTR is shown above the top entry. Mismatches are highlighted. The symbols for polyamides are: open, filled and slashed circles, Im, Py and Bz rings, respectively; curved lines,  $\gamma$ -aminobutyric acid; diamonds,  $\beta$ -alanine; dimethylaminopropylamine tail and amino group of (*R*)-2,4-diaminobutyric acid, plus signs.

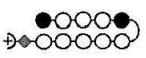
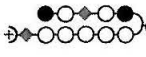
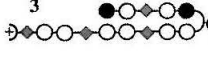




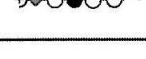
The most conserved feature of all retroviral termini is the CA/GT dinucleotide pair, invariably found precisely at the site of joining to the host DNA. Sequences internal to the CA, extending up to 15 bp from the termini, have a more limited effect on IN activity.<sup>43-46</sup> All polyamides generated target the CA dinucleotide and vary in the extent

of the surrounded sequences to which they bind (Fig. 1B). Polyamides **1-3** were designed to bind a core sequence from positions 2-9 of the LTR. The central Py/Py pair of **1** was varied to a  $\beta$ /Py (**2**) or  $\beta/\beta$  pair (**3**). Polyamide **1** contains the standard  $\gamma$ -turn, whereas hairpins **2** and **3** contain the charged chiral analog,  $(R)^{H_2N}\gamma$ . These latter compounds are doubly charged. Polyamide **3**, which contains a C-terminal extension of  $\beta$ -Py-Py, is designed to bind a larger site, placing a Py ring at position 1 of the LTR. Polyamide **4** represents a single base pair mismatch for the 2-9 region. Another control compound, polyamide **8**, is a double base pair mismatch for the LTR sequence. The recent development the A•T and T•A selective N-terminal Bz residue allowed for the targeting of additional sequences, which place the core of the hairpins closer to the terminus of the LTR (**5-7**). Polyamides **5** and **7** cover positions 1-7, and polyamide **6** covers positions 2-8. While polyamides **1-4** place a Im/Py pair and un-paired  $\beta$ -alanine against the CA dinucleotide, polyamides **5-7** place two central ring pairings against the CA dinucleotide, which may serve as a greater steric impediment to interaction with the integrase.

### **High-affinity binding of the polyamides to the viral LTR sequence.**

Quantitative DNase I footprinting titration experiments<sup>40</sup> were performed to determine the equilibrium association constant ( $K_a$ ) of the polyamides with the M-MuLV LTR sequence (Table 2.1, see also Chapter 3). Individual polyamides were incubated with a <sup>32</sup>P labeled DNA fragment containing an internal copy of the LTR, followed by DNase I digestion. Note that this procedure requires an excess of polyamide, and the amount of DNA used in this procedure is on the order of 5 pM.<sup>40</sup> Representative DNase I footprinting gels for polyamides **2** and **6** are shown in Figure 2.2. Polyamides **5-7**, which are constitutional isomers, varying only by the position of a single nitrogen atom within the molecule, are readily differentiated by the LTR sequence (Table 2.1). Among the polyamides tested, polyamide **5** has the highest equilibrium association constant

**Table 1.** Equilibrium association constants for polyamides with internal copy of MuLV LTR.

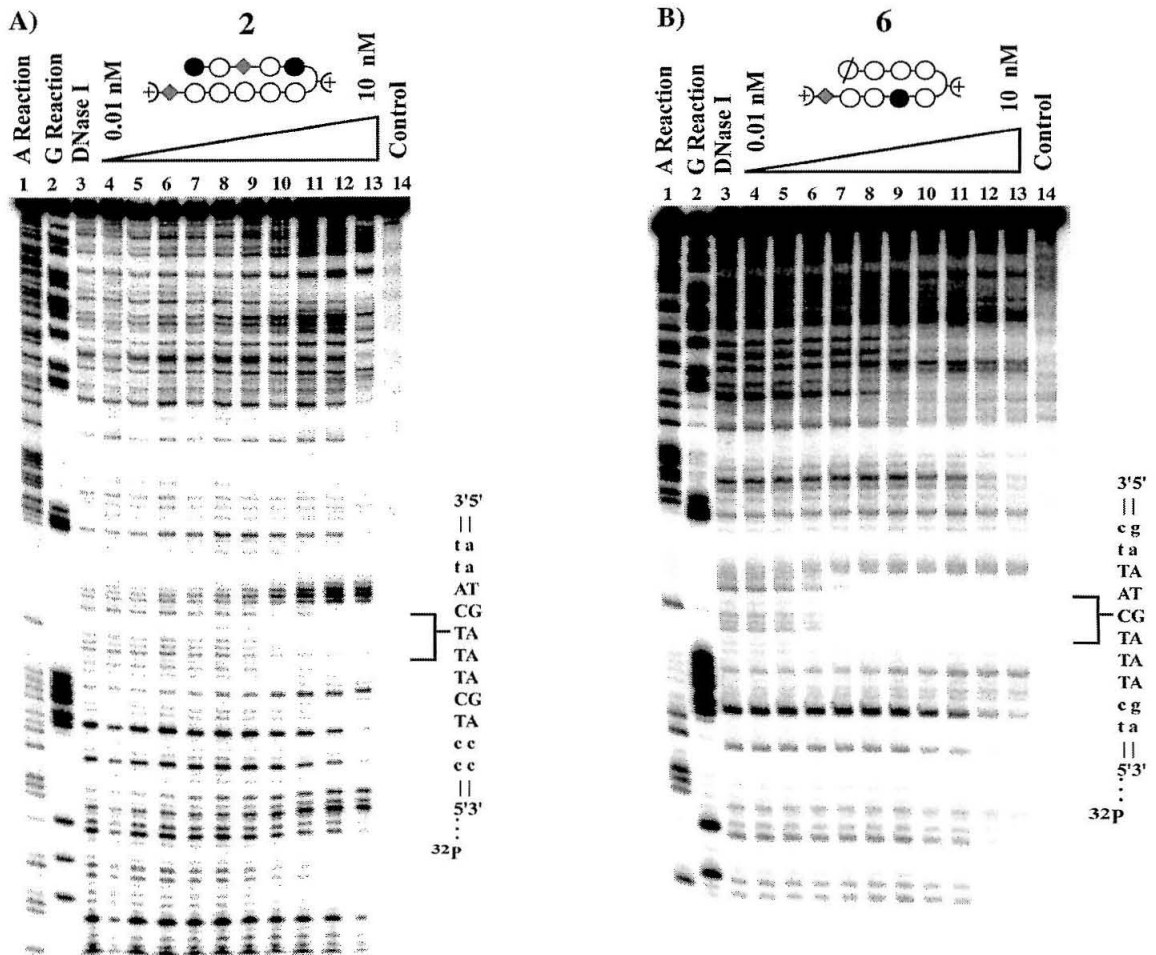
Polyamide	Sequence	K <sub>a</sub> (M <sup>-1</sup> )
1 	5' -GGGAATGAAAGACC-3' <u>          </u>	2.2 x 10 <sup>9</sup>
2 	5' -GGGAATGAAAGACC-3' <u>          </u>	3.6 x 10 <sup>9</sup>
3 	5' -GGGAATGAAAGACC-3' <u>          </u>	2.7 x 10 <sup>9</sup>
4 	5' -GGGAATGAAAGACC-3' <u>          </u>	2.1 x 10 <sup>7</sup>
5 	5' -GGGAATGAAAGACC-3' <u>          </u>	4.0 x 10 <sup>10</sup>
6 	5' -GGGAATGAAAGACC-3' <u>          </u>	1.7 x 10 <sup>10</sup>
7 	5' -GGGAATGAAAGACC-3' <u>          </u>	3.0 x 10 <sup>9</sup>
8 	5' -GGGAATGAAAGACC-3' <u>          </u>	≤5 x 10 <sup>6</sup>

**Table 2.1.** Values reported are the mean values obtained from three DNase I footprint titration experiments. The binding site for each polyamide is underlined. Mismatches are shown in bold. The assays were carried out at 22° C, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

(K<sub>a</sub> = 4 x 10<sup>10</sup> M<sup>-1</sup>), while the double base pair mismatch **8** has the lowest binding constant (K<sub>a</sub> ≤ 5 x 10<sup>6</sup> M<sup>-1</sup>).

The similar affinity (K<sub>a</sub> = 1.9 – 3.6 x 10<sup>9</sup> M<sup>-1</sup>) of polyamides **1-3** is of interest. Polyamide **1** appears to be particularly well matched for the LTR sequence, given its relatively high affinity, and observations that substitution with a central β/β pair and/or the positively charged (R)<sup>H<sub>2</sub>N</sup>γ-turn failed to increase its affinity (data not shown). Additionally, substitution with a central β/Py pair yielded a drop in affinity, which was recovered by substitution with (R)<sup>H<sub>2</sub>N</sup>γ (polyamide **2**). With the internal copy of the LTR sequence used for these experiments, the tail of the extended polyamide **3** is placed against G•C base pairs, which represents a mismatch that is not present in the LTR itself. It is important to note that, since the DNase I footprinting titrations were performed on an internal copy of the LTR sequence, the affinity constants generated are only a model for

Figure 2.2



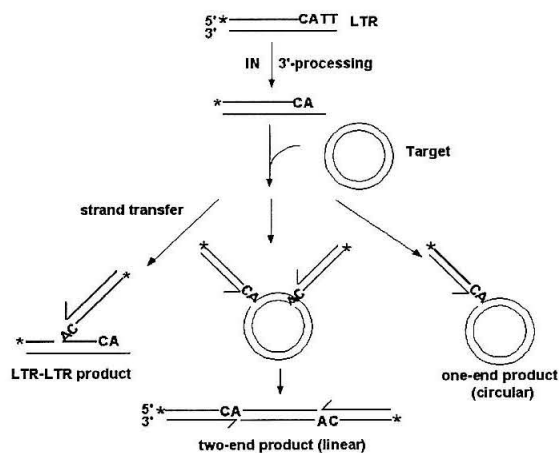
**Figure 2.2.** (A) Quantitative DNase I footprint titration experiment with polyamide **2** on a 5'  $^{32}\text{P}$  labeled 183 base pair DNA fragment containing an internal copy of the M-MuLV LTR: lane 1, A reaction; lane 2, G reaction; lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide; lane 14, intact DNA. All reactions contained 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>, and were performed at 22° C. (B) Quantitative DNase I footprint titration experiment with polyamide **6** on a 5'  $^{32}\text{P}$  labeled 183 base pair DNA fragment containing an internal copy of the M-MuLV LTR: lane 1, A reaction; lane 2, G reaction; lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide; lane 14, intact DNA. All reactions contained 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>, and were performed at 22° C.

binding to the M-MuLV LTR termini. The binding of polyamides **5-7** was of particular interest, given the proximity of their cognate sites to the absolute end of the LTR. DNA melting experiments using oligonucleotides with terminal or internal cognate sites were



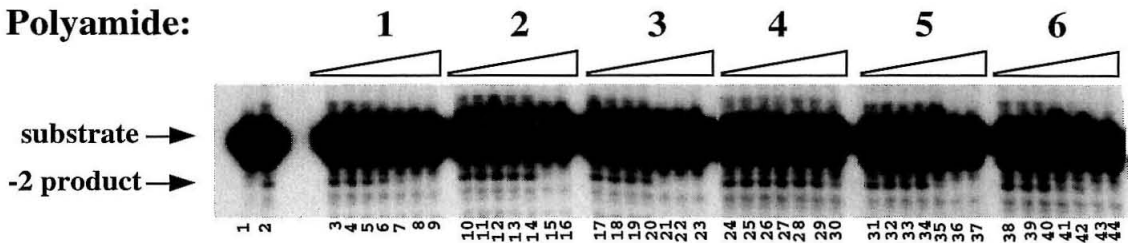
performed to address this issue.<sup>47</sup> Comparisons of melting curves with and without polyamides **5-7** showed that the  $T_m$  shifts induced by these polyamides were similar for both sets of oligonucleotides, although slightly lower in each case for the terminal cognate site (see Appendix). In all cases, the  $T_m$  shifts for both sets of oligonucleotides with polyamides **5-7** were considerably greater than the  $T_m$  shifts induced by a mismatch polyamide. These results indicate that polyamides can bind with high affinity to the absolute ends of DNA fragments. However, since the  $T_m$  shift for the terminal cognate site was slightly lower than the internal cognate site for polyamides **5-7**, the affinity constants determined by DNase I footprinting for these compounds (Table 2.1) should be considered as upper limits for their affinity to the M-MuLV LTR termini.

**Figure 2.3**



**Figure 2.3.** Schematic illustration of *in vitro* integration assays. IN first removes the "TT" dinucleotide from the 5' <sup>32</sup>P labeled blunt LTR duplex, leaving a single-strand tail. Then IN integrates the cleaved strand into the target: either the LTR itself in the strand transfer assay, or the plasmid DNA as an exogenous substrate. The major products of integration are as listed.

**Inhibition of the 3' processing.** The ability of the polyamides to inhibit integration was tested in *in vitro* integration assays. Figure 2.3 schematically outlines the *in vitro* assays utilized in these studies. In the 3' processing assay, a 20 base-pair (bp) blunt-end oligonucleotide duplex mimicking the viral LTR end was used as the substrate

**Figure 2.4****Polyamide:**

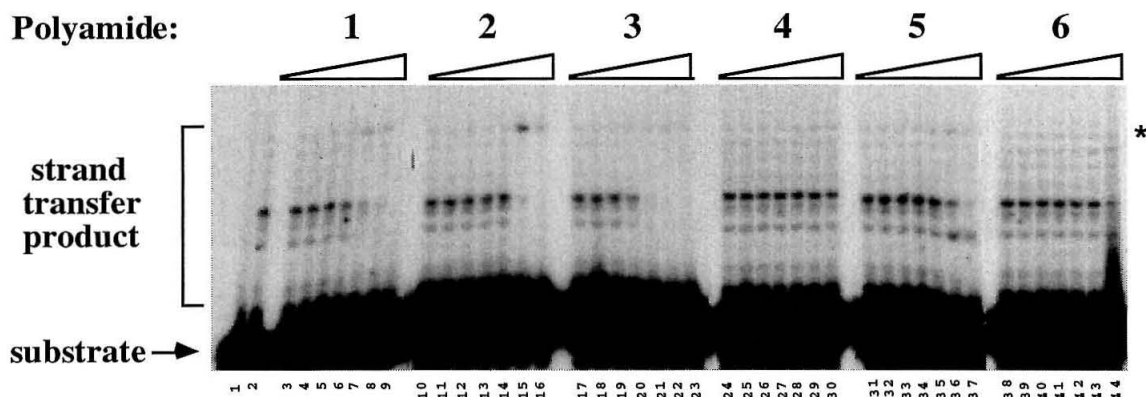
**Figure 2.4.** Inhibition of 3' processing reactions by the polyamides. 3' processing reactions were performed as described in the Materials and Methods. The position of the -2 3' processing product is indicated at the left. Lane 1, no protein control; lane 2, M-MuLV IN alone. Lanes 3-9, IN plus polyamide 1; lanes 10-16, IN plus polyamide 2; lanes 17-23, IN plus polyamide 3; lanes 24-30, IN plus polyamide 4; lanes 31-37, IN plus polyamide 5; and lanes 38-44, IN plus polyamide 6. The concentrations of the polyamide used are (left to right): 0.4, 16, 40, 100, 256, 640 and 1600 nM.

(Figure 2.3, top line). The strand containing the conserved CA dinucleotide was 5'  $^{32}\text{P}$  labeled. IN dependent 3' processing of the dinucleotide yields an 18 base product, which is monitored on a denaturing acrylamide gel. It is of note that the concentration of DNA substrate in these and the following integration assays is significantly greater (33 nM) than the concentration of DNA in the DNase I footprinting titrations. Figure 2.7 summarizes the effects of the polyamides on the 3' processing reaction. The results of an exemplary 3' processing assay with selective polyamides is shown in Figure 2.4. The eight polyamides could be divided into three classes, according to their inhibition efficiencies (Figure 7). The first class includes polyamides **1**, **3**, **5**, and **6** all having an  $\text{IC}_{50}$  under 250 nM. The second class includes polyamides **2** and **7** with an  $\text{IC}_{50}$  between 450 and 650 nM. Polyamides **4** and **8** belong to the third class, which has an  $\text{IC}_{50}$  at or above 1400 nM. The polyamides are designed to bind with sequence specificity. The inhibitory effect of the polyamides correlates with the ability to efficiently bind the LTR sequences and is therefore not due to nonspecific interactions with the IN protein. A single mismatch within the polyamide was found to decrease the  $K_a$  value by 100-fold and showed little evidence of inhibition of 3' processing and strand transfer. This is

exemplified by polyamide 4 (Figure 2.4 and data not shown), with a  $K_a$  of  $2.1 \times 10^7 \text{ M}^{-1}$ , and an  $\text{IC}_{50}$  value at or greater than 1600 nM.

**Figure 2.5**

**Polyamide:**



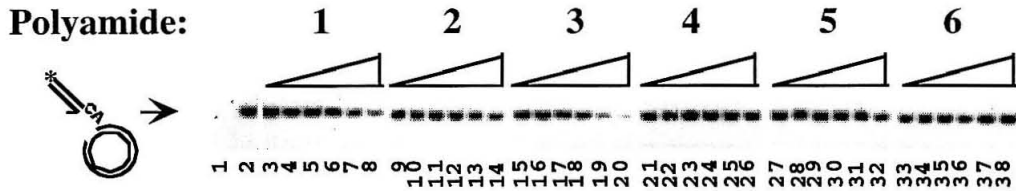
**Figure 2.5.** Inhibition of strand transfer reactions by the polyamides. Strand transfer assays, based on oligonucleotides substrates were performed as described in the Materials and Methods. Strand transfer yields  $^{32}\text{P}$ -labeled products larger than the input substrate, indicated at left of panel. Lane 1, no protein control; lane 2, M-MuLV IN alone. Lanes 3-9, IN plus polyamide 1; lanes 10-16, IN plus polyamide 2; lanes 17-23, IN plus polyamide 3; lanes 24-30, IN plus polyamide 4; lanes 31-37, IN plus polyamide 5; and lanes 38-44, IN plus polyamide 6. The concentrations of the polyamide used are (left to right): 0.4, 16, 40, 100, 256, 640, and 1600 nM. The asterisk indicates the formation of the enhanced integration product.

**Inhibition of strand transfer.** The polyamides were tested for inhibition of strand transfer. In this assay, a precleaved substrate (Figure 2.3, second row) was used instead of the blunt-end substrate, with integration occurring into a second double-stranded oligonucleotide (Figure 2.3, left arrow). Strand transfer is highly efficient using a precleaved substrate,<sup>48</sup> and allows for the separation of the two steps in the analysis. Similar to 3' processing, the polyamides can be separated into three classes based on their  $\text{IC}_{50}$  for the strand transfer reactions (Figure 2.7). An exemplary strand transfer assay using selective polyamides is shown in Figure 2.5. The first class consists of 1 and 3, with the  $\text{IC}_{50}$  under 250 nM. Moderate inhibition is found with the second class, including 2 and 5 with  $\text{IC}_{50}$  at 450 and 1100 nM, respectively. The third class includes 4 and 6-8. The  $\text{IC}_{50}$  of the third class are at or above 1400 nM.

Although polyamides **5** and **6** were among the most effective inhibitors of 3' processing, they showed little effect on strand transfer. Polyamides **1** and **3**, though, are highly effective at inhibition of both the 3' processing and strand transfer reactions. This raises the interesting possibility that distinct regions of the viral termini are required for 3' processing and strand transfer. Additionally, the absence of the TT dinucleotide in the strand transfer substrate will affect the binding of polyamides, particularly **5-7**, whose cognate sites are closest to the end of the viral DNA. Each of these polyamides was less effective at inhibiting strand transfer than 3'-processing, whereas polyamides **1-3** showed the same efficacy in both assays (Figure 2.7). This further emphasizes the importance of cognate site recognition in integrase inhibition.

In the oligonucleotide based assay, the LTR substrate functions both as the viral and target DNA. Previous analysis of binding site selection indicated that the major groove was the main determinant within the target DNA for IN docking.<sup>49-53</sup> Therefore it was believed that binding of the polyamides to the target DNA should not influence target site selection. Experimentally, variations in target site selection were observed. In the presence of high concentration of polyamide **2** (640 and 1600 nM, Figure 2.5, lanes 15 and 16), loss of integration at the predominant sites was paralleled with enhanced production of a single product of slow electrophoretic mobility (marked by the asterisk). This product was also observed at saturating levels of polyamide **3** (Figure 2.5, lanes 21-23). Formation of this large product is a result of integration at one end of the target DNA. Target site selection using polyamide **5** was also altered, with preference for the faster mobility products at the highest polyamide concentrations (Figure 2.5, lanes 36-37).

**Inhibition of strand-transfer into an exogenous target.** To separate the effects of polyamide binding to the viral versus target site, an alternative *in vitro* integration assay was used (Figure 2.3), in which a precleaved LTR oligonucleotide donor was

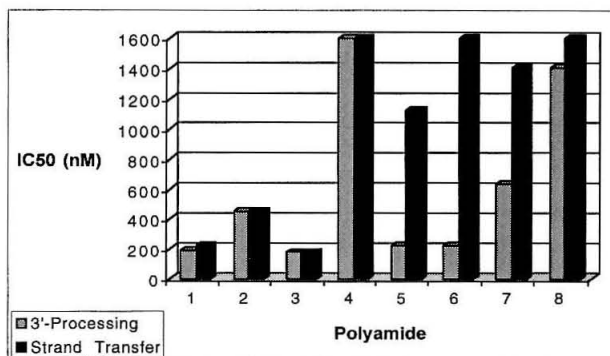
**Figure 6**

**Figure 6.** Polyamides inhibit *in vitro* strand transfer into an exogenous substrate.  $^{32}\text{P}$ -labeled LTR substrate was analyzed for integration into an unlabeled plasmid DNA. The migration on an agarose gel of the single-end lariat integration product is indicated at the left of the panel. Lane 1, no protein control; lane 2, M-MuLV IN alone. Lanes 3-8, IN plus polyamide 1; lanes 9-14, IN plus polyamide 2; lanes 15-20, IN plus polyamide 3; lanes 21-26, IN plus polyamide 4; lanes 27-32, IN plus polyamide 5; and lanes 33-38, IN plus polyamide 6. The concentrations of the polyamide used are (left to right): 16, 40, 100, 256, 640 and 1600 nM.

integrated into a nonspecific circular target plasmid. The target DNA does not encode the LTR sequence and therefore only the effect of the polyamide on the LTR is examined. Integration of the LTR into the target plasmid would result in two products: circular "lariat like" DNA from integration of one LTR end, and linear DNA from integration of two LTR ends in coordination (Figure 2.3, right and center arrows, respectively). The integration products into the plasmid DNA are visualized after separation on an agarose gel. Using WT MuLV IN, the predominant product was the lariat product (Figure 2.6). The general efficiency of the polyamides was as found with the oligonucleotide based assay. Examples of polyamides in which no inhibition was observed include polyamides 4 and 6 (Figure 2.6, lanes 21-26 and lanes 33-38, respectively). Inhibition by polyamide 2 and 5 (Figure 2.6, lanes 14 and 32, respectively) was observed at the highest concentration tested (1600 nM). Polyamides 1 and 3 (Figure 2.6, lanes 3-8 and 15-20, respectively) were effective in inhibiting strand transfer between 250 and 640 nM. Interestingly, polyamide 3 yielded reproducibly the lowest  $\text{IC}_{50}$ , indicating recognition of the LTR sequence in the presence of excess nonspecific target DNA. Longer exposure of the gels for the two-end linear products paralleled the results for the one-end integration (data not shown). The similar profiles observed in

both the two-end integration assay and the strand transfer assay indicate that the two are mechanistically the same.

**Figure 2.7**



**Figure 2.7.**  $IC_{50}$  of polyamides on oligonucleotide-based M-MuLV integrase activities. The inhibitory effect of each polyamide (0.5-1600 nM) was titrated in either the 3' processing (grey) or strand transfer (black) assays. Individual bands of reaction products were quantified using Scion Image 1.62C. For each assay, reaction products of the MuLV IN catalyzed in the absence of polyamides was defined as 100% activity. Equivalent fractions from lanes in the absence of enzyme were taken as background and subtracted from each reaction.  $IC_{50}$  values were determined as the concentration of each polyamide that inhibited IN activities by 50%.

## Discussion

With the appearance of HIV strains resistant to both the protease and the reverse transcriptase inhibitors, the need to develop potent inhibitors against new targets including the integrase is more urgent.<sup>54</sup> The increased interest for inhibitors which target the integrase has provided lead compounds,<sup>19,54</sup> although no clinically effective inhibitor is available to date.

In this report, a series of polyamides were designed and tested as inhibitors for integration by the M-MuLV IN. Previous studies have shown that DNA-binding agents can effectively inhibit retroviral integration.<sup>21,22,25,55,56</sup> However, all the compounds have very limited sequence specificity and can bind to a very large number of potential DNA sites. The polyamides developed in this report bind specifically to the minor groove of sequences required for retroviral integration, those at or flanking the conserved CA

dinucleotide. Polyamides were identified with subnanomolar binding constants. The polyamides were tested in three different integration assays in cell free experiments. The inhibition of retroviral integration observed was due to binding of the polyamides to the LTR termini. Polyamides with mismatch recognition of the MuLV LTR were much less efficient inhibitors. This is exemplified by polyamides **4** and **8**, which both bind to the M-MuLV LTR with  $K_a$  values of no greater than  $2 \times 10^7 \text{ M}^{-1}$  and have  $\text{IC}_{50}$  at or greater than 1400 nM. The inhibition observed, therefore, is not due to a nonspecific binding of the polyamides to DNA or to the IN protein. A clear dependence on cognate site recognition is demonstrated by the discrepancy between the 3' processing and strand transfer results for polyamides **5-7**. The binding site of these polyamides is affected by the removal of the terminal TT dinucleotide, and thus the cognate polyamides showed reduced strand transfer inhibition as compared to 3' processing, where the cognate site is intact.

Despite the importance of cognate site recognition, the  $K_a$  values determined by DNase I footprinting with an internal copy of the M-MuLV LTR were not direct predictors of the efficacy of integrase inhibition. In general, all of the polyamides containing an  $\text{IC}_{50}$  lower than 640 nM, had binding constants in the nanomolar range or lower. However, this correlation is not reciprocal. Polyamides with perfect sequence matches show subtle variations with the correlation of the  $K_a$  values and the effective  $\text{IC}_{50}$ . Polyamides **2** and **7** have  $K_a$  values near  $3.0 \times 10^9 \text{ M}^{-1}$ , yet are not as effective at inhibiting 3' processing as polyamides **1** and **3**, which have association constants in the same range. As noted, given the proximity of its binding sites to the end of the LTR, the  $K_a$  value for **7** should be considered an upper limit. This may also explain why the efficacy of polyamides **5** and **6** does not exceed that of polyamides **1** and **3**. The lower activity of **2** is more difficult to explain, given its moderately higher affinity than **1**.

The similarity efficiency of **1** and **3** is of interest. When footprinted against the internal copy of the LTR, the tail of polyamide **3** was placed over G•C base pairs, which



is considered a mismatch.<sup>57</sup> Under these conditions, its association constant was quite close to polyamide **1**. These G•C pairs are not present in the 3'-processing substrate LTR, where the tail of **3** extends beyond the terminal base pair. It is thus possible that placing the  $\beta$ -Dp tail of this extended-hairpin polyamide beyond the terminal base pair has energetic consequences similar to placing it against G•C base pairs. However, if **1** and **3** do have the same affinity, one might expect that **3** would be the better inhibitor of 3'-processing, due to its C-terminal extension, which places Py residues against the bases to be excised. Given that these bases are absent in the strand transfer substrate, it is also interesting that **3** maintains its activity in this assay, while the activity of **5-7** drops. In fact, in the presence of excess unlabeled DNA in the exogenous target assay, polyamide **3** was the most effective strand transfer inhibitor. Since the hairpin portion of **3** is internal to the excised nucleotides, it can be speculated that **3** maintains high affinity binding through its hairpin portion. Perhaps the long C-terminal tail provides a steric impediment to strand transfer, even in the absence of cognate base-paired DNA.

It is also significant that all of the effective strand transfer inhibitors bind positions 7 through 9 of the M-MuLV LTR. This is consistent with the observations of Balakrishnan and others in the respective HTLV (human T-lymphotropic virus) and HIV-1 systems of the importance of these residues in related viral LTRs.<sup>45,46,58</sup> Positions 7-9 are bound by either the tail or turn of polyamides **5-7**, or are outside of the binding site of these polyamides. The turn and tail residues are not considered to be as effective steric blockades of the minor groove as ring pairings. Thus, reduced blockage of positions 7-9 relative to polyamides **1-3** correlates with reduced strand transfer inhibition, although this analysis is complicated by the expectation that polyamides **5-7** will have reduced affinity for the strand transfer substrate. It would be interesting to test the ability of a polyamide that binds internally to positions 7-9 to act as a strand transfer inhibitor.

The integration process itself is a dynamic process. Current models predict conformational changes accompanying cleavage of the viral DNA terminus.<sup>30</sup>



Differential cross-linking efficiencies to the HIV-1 LTR sequences were observed for substrates for 3' processing, strand transfer, and disintegration.<sup>30</sup> In addition, the binding of the diketo acid inhibitor L-731,988, required the assembly of the IN protein with the viral end. L-731,988 competes with the target DNA, indicating the target binding site is generated after the binding of the viral LTR. These compounds selectively inhibited strand transfer, without any effect on 3' processing.<sup>18</sup> Cross-linking studies with HIV-1 IN indicate that the C-terminal domain of one IN monomer acts in *trans* with the catalytic core of another monomer for each viral end.<sup>30</sup> This in *trans* assembly may involve the sequences upstream of the viral LTRs, within the region bound by polyamides **1** and **3**. Similarly, structural studies of Tn5 transposase have demonstrated that both the internal and the terminal regions of the transposon DNA are contacted by the transposase.<sup>59</sup> This conformational change conceptually follows the assembly and transport of the preintegrative complexes. *In vivo*, 3' processing is temporally and spatially independent of strand transfer.<sup>60</sup> 3' processing occurs in the cytoplasm for MuLV, whereas strand transfer occurs in the nucleus. Initial binding of the viral DNA requires the recognition of the linear viral terminus, including the CA dinucleotide. The requirements of protein domains for 3' processing are stringent; for MuLV IN all domains, including the N-terminal HHCC region, are required.<sup>7,61,62</sup>

The substrate for 3' processing is a blunt-end double strand DNA. Fraying or unwinding of the termini has been proposed to be a requisite step in the 3' processing reaction.<sup>63,64</sup> Several of the polyamides target the terminal dinucleotides. It was possible that the binding of polyamide might impede the unwinding step. However, no preferential inhibition of 3' processing was observed for the polyamides which bound to the terminal TT sequence. Polyamides which bind to the terminal nucleotides may, however, sterically hinder subsequent assembly of a strand transfer complex. Polyamide **3** recognizes the largest binding site, from position 1 to position 9 of the LTR, including a single-stranded tail covering the terminal TT dinucleotide. As discussed, this may

contribute to the overall effectiveness of inhibition by polyamide **3** in both the 3' processing and strand transfer assays.

The variation in target site selection in the presence of polyamides is of interest. Previous studies indicated for HIV-1 IN, target sites were preferred within nucleosomes with wide major grooves (50). Altered target site selection induced by the polyamides binding to the minor groove can indicate either that minor-groove contacts are important or that structure of the major groove can be altered by binding of polyamides to the minor groove. Structural evidence suggests that hairpin polyamides bind a wide minor groove, with requisite narrowing of the major groove (65-68). Therefore, it is possible that presence of the polyamides can have an indirect effect on the interaction of the pre-integrative complex within the major groove. It is also expected that binding of polyamides in the minor groove will reduce the flexibility of a given stretch of DNA, reducing access to DNA conformations that could be important in the dynamic process of target site selection.

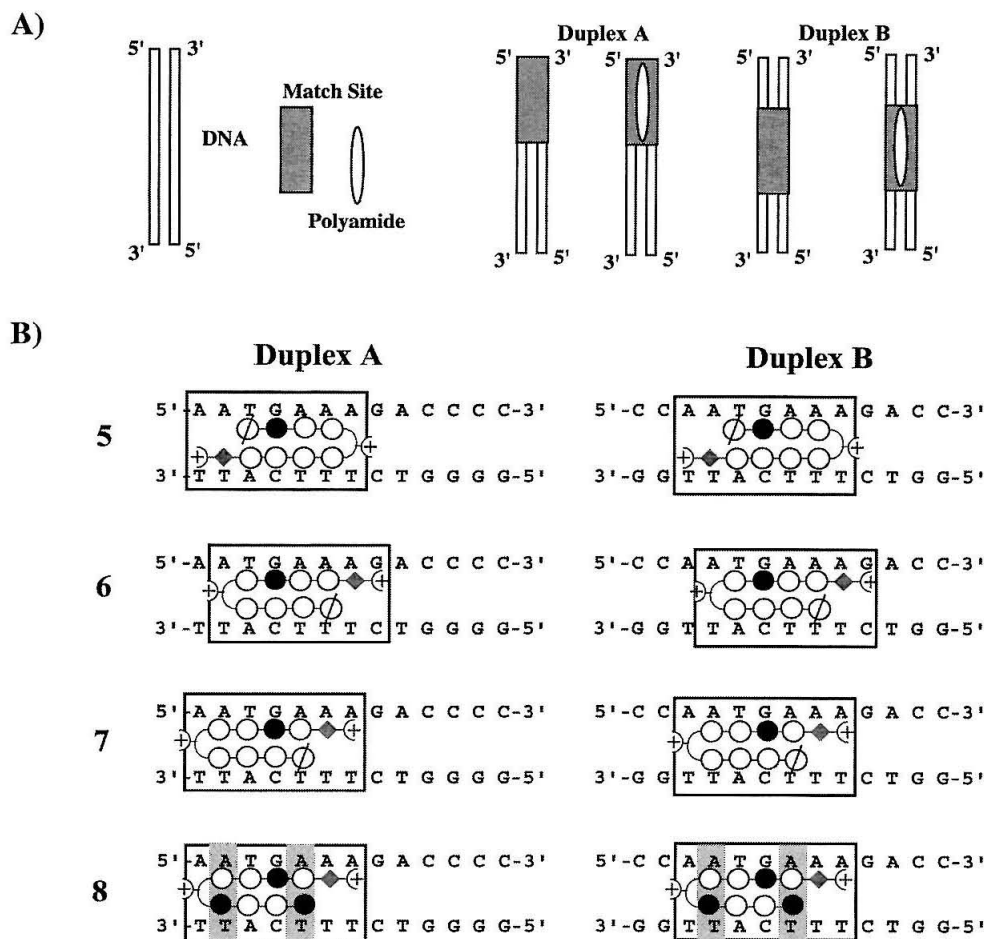
These studies represent the first generation of polyamides which inhibit MuLV integration. Four potent inhibitors were developed, two which inhibit 3'-processing (**5** and **6**), and two which inhibit both 3' processing and strand transfer (**1** and **3**). These different classes of inhibitors could be useful tools for unraveling the complex mechanism of integration. Both sets of compounds are effective at a less than 10-fold excess over substrate DNA and less than half an equivalent of integrase protein. Next generation polyamides can incorporate hydroxypyrrole<sup>68</sup> and hydroxybenzamide<sup>38</sup> residues, which break the A•T/T•A degeneracy of the Py/Py and Bz/Py pairs, allowing for greater sequence specificity. Strategies other than increasing the binding efficiency could make the inhibitors more potent. Alterations in the structure of the polyamides could change the conformation of the LTR and/or interfere with the binding of the IN. Modified polyamides which may be able to clamp the terminal TT dinucleotide by alkylation<sup>69,70</sup> may be envisioned. Certainly, a crystal structure of an IN-LTR complex

would aid the design of polyamide inhibitors. Alternatively, the polyamides can serve as the matrix to be coupled with additional inhibitors, for example, compounds that bind the enzyme active site.<sup>71</sup> Delivery of these compound inhibitors to the enzyme would be facilitated by the binding of the polyamides to the viral DNA.

The HIV-1 IN and M-MuLV IN share the same mechanism of integration. It is very likely that the same strategy of designing polyamide inhibitors can be applied to the HIV-1 IN. Studies are currently underway to test polyamides with the HIV-1 IN/LTR system. Given the previously demonstrated efficacy of polyamides targeted to HIV-1 in human cells (33), this approach may ultimately yield potent *in vivo* HIV-1 IN inhibitors. The modular nature of polyamide recognition and synthesis should allow for rapid and rational changes to the inhibitors in response to viral mutations.

## Appendix: DNA Melting Experiments

A set of DNA melting experiments was designed to test the ability of hairpin polyamides to bind to the absolute ends of DNA fragments. Duplex **A** represents the terminal 13 bp of the M-MuLV LTR, 5'-AATGAAAGACCCC-3' (Figure 2.8). Duplex **B** has the same nucleotide composition, with two base pairs moved from the 3' to the 5' end of the duplex, which shifts the binding site of the polyamides away from the end of the duplex. Polyamide **5** places Dp against position 1 of duplex **A** and position 3 of duplex **B**. Polyamide **6** places the turn  $((R)^{H_2N}\gamma)$  against position 2 of duplex **A** and position 4 of duplex **B**. Polyamide **7** places the turn  $((R)^{H_2N}\gamma)$  against position 1 of duplex **A** and position 3 of duplex **B**. Polyamide **8** is a double base pair mismatch for both duplexes. Melting curves were obtained by standard UV-vis spectroscopy (80° C—15° C, 1° C/minute, scan at 260 nM every 30 seconds) for both duplexes (2 μM) in the absence and presence of 2 μM polyamide (**5-8**). These experiments were performed using buffered conditions (see legend, Table 2.2) designed by Pilch *et al.* to be similar to those used for DNase I footprinting titrations.<sup>47</sup>

**Figure 2.8**

**Figure 2.8** A) Cartoon of the DNA melting experiments. The same binding site is placed at the end and middle of two short DNA duplexes. B) Sequences of the duplexes A and B with ball-and-stick models for the polyamides showing the binding site in a box and mismatches in gray.

All polyamides tested increased the stability of both duplexes as measured by an increase in the melting temperature ( $T_m$ ) of the duplex in the presence of the polyamide (Table 2.2). As expected, for duplex B with the internal match site, the  $\Delta T_m$  for match polyamides was significantly higher than the mismatch polyamide and the order of increasing  $\Delta T_m$  matched the order of increasing  $K_a$  as measured by DNase I footprinting ( $8 < 7 < 6 < 5$ ). The same order was observed for duplex A, and the  $\Delta T_m$  of the match polyamides were all significantly higher than the mismatch. This indicates that

**Table 2.2.** DNA melting with internal and terminal polyamide binding sites.<sup>a</sup>

	K <sub>a</sub> (DNase I) <sup>b</sup> (M <sup>-1</sup> )	T <sub>m</sub> Duplex <b>B</b> <sup>c</sup> (°C)	T <sub>m</sub> Duplex <b>A</b> <sup>d</sup> (°C)	ΔT <sub>m</sub> <b>B</b> <sup>c</sup> (°C)	ΔT <sub>m</sub> <b>A</b> <sup>d</sup> (°C)	ΔΔT <sub>m</sub> <sup>e</sup> (°C)	% ΔΔT <sub>m</sub> / ΔT <sub>m</sub> <b>B</b>
---	---	45.9	46.9	---	---	---	---
<b>5</b>	4.0x10 <sup>10</sup>	69.9	68.0	24.0	21.1	2.9	12%
<b>6</b>	1.7x10 <sup>10</sup>	67.5	66.4	21.6	19.5	2.1	10%
<b>7</b>	3.0x10 <sup>9</sup>	63.0	62.5	17.1	15.6	1.5	9%
<b>8</b>	≤5x10 <sup>6</sup>	54.8	53.9	8.9	7.0	1.9	21%

**Table 2.2** <sup>a</sup>Melting curves (80° C—15° C, 1° C/minute) were obtained in 10 mM sodium cacodylate, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>. 2 μM polyamide and 2 μM DNA duplexes **A** or **B** were used in each experiment. <sup>b</sup>See Table 2.1 and Experimental Procedures. <sup>c</sup>ΔT<sub>m</sub> **B** is defined in the difference in melting temperature of duplex **B** with the addition of polyamide. <sup>d</sup>ΔT<sub>m</sub> **A** is defined in the difference in melting temperature of duplex **A** with the addition of polyamide. <sup>e</sup>ΔΔT<sub>m</sub> is defined as ΔT<sub>m</sub> **B** - ΔT<sub>m</sub> **A**.

polyamides can bind specifically to a match site at the absolute end of a DNA fragment with high affinity, although in each case ΔT<sub>m</sub> **A** was slightly lower than ΔT<sub>m</sub> **B**. For each of the match polyamides, the ΔΔT<sub>m</sub> was approximately 10% of ΔT<sub>m</sub> **B**. A rough calculation using the relationship  $\Delta T_m \propto \ln K_a$  suggests that polyamides **5-7** have a 3-fold lower affinity for the terminal match site as compared to an internal match site (assuming that polyamides **5-7** bind the match site on duplex **B** with similar affinity to the same match site on the larger fragment used for DNase I footprinting). It is interesting to note that the penalty for binding to the end of duplex **A** is essentially the same whether the turn or the tail is placed against the first base pair, or the turn is placed against the second base pair. The lack of difference between the terminal two base pairs is likely due to fraying of terminal A,T base pairs. Duplex **A**, which represents the natural M-MuLV termini, may be especially prone to fraying of the terminal base pairs, as this is likely an important structural element for integrase recognition.<sup>63</sup>

## Acknowledgements

This work was supported by NIH Grants CA76545 to M. J. R. and GM57148 to P.B.D. We are grateful to the Ralph M. Parsons Foundation for a predoctoral fellowship to J.M.B. We thank Don Crothers (Yale) for helpful discussions concerning the DNA melting experiments and the Barton lab for the use of the UV-vis spectrometer equipped for these experiments.

## References

- (1) Brown, P.O. *Integration*; Cold Spring Harbor Laboratory Press: Plainview, New York, 1997.
- (2) Hindmarsh, P.; Leis, J. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 836-843, table of contents.
- (3) Katzman, M.; Katz, R. A.; Skalka, A. M.; Leis, J. *J. Virol.* **1989**, *63*, 5319-5327.
- (4) Craigie, R.; Fujiwara, T.; Bushman, F. *Cell* **1990**, *62*, 829-837.
- (5) Sherman, P. A.; Fyfe, J. A. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 5119-5123.
- (6) Singh, I. R.; Crowley, R. A.; Brown, P. O. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 1304-1309.
- (7) Yang, F.; Roth, M. J. *J. Virol.* **2001**, *75*, 9561-9570.
- (8) Neamati, N.; Hong, H.; Owen, J. M.; Sunder, S.; Winslow, H. E.; Christensen, J. L.; Zhao, H.; Burke, T. R., Jr.; Milne, G. W.; Pommier, Y. *J. Med. Chem.* **1998**, *41*, 3202-3209.
- (9) Zouhiri, F.; Mouscadet, J. F.; Mekouar, K.; Desmaele, D.; Savoure, D.; Leh, H.; Subra, F.; Le Bret, M.; Auclair, C.; d'Angelo, J. *J. Med. Chem.* **2000**, *43*, 1533-1540.
- (10) Drake, R. R.; Neamati, N.; Hong, H.; Pilon, A. A.; Sunthankar, P.; Hume, S. D.; Milne, G. W.; Pommier, Y. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4170-4175.

- (11) Taktakishvili, M.; Neamati, N.; Pommier, Y.; Nair, V. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 249-251.
- (12) Jing, N.; Hogan, M. E. *J. Biol. Chem.* **1998**, *273*, 34992-34999.
- (13) Mazumder, A.; Neamati, N.; Ojwang, J. O.; Sunder, S.; Rando, R. F.; Pommier, Y. *Biochemistry* **1996**, *35*, 13762-13771.
- (14) Rando, R. F.; Ojwang, J.; Elbaggari, A.; Reyes, G. R.; Tinder, R.; McGrath, M. S.; Hogan, M. E. *J. Biol. Chem.* **1995**, *270*, 1754-1760.
- (15) Jing, N.; Marchand, C.; Liu, J.; Mitra, R.; Hogan, M. E.; Pommier, Y. *J. Biol. Chem.* **2000**, *275*, 21460-21467.
- (16) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13040-13043.
- (17) Lubkowski, J.; Yang, F.; Alexandratos, J.; Wlodawer, A.; Zhao, H.; Burke, T. R. Jr.; Neamati, N.; Pommier, Y.; Merkel, G.; Skalka, A. M. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4831-4836.
- (18) Espeseth, A. S.; Felock, P.; Wolfe, A.; Witmer, M.; Grobler, J.; Anthony, N.; Egbertson, M.; Melamed, J. Y.; Young, S.; Hamill, T.; Cole, J. L.; Hazuda, D. J. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11244-11249.
- (19) Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. *Science* **2000**, *287*, 646-650.
- (20) Fesen, M. R.; Pommier, Y.; Leteurtre, F.; Hiroguchi, S.; Yung, J.; Kohn, K. W. *Biochem. Pharmacol.* **1994**, *48*, 595-608.
- (21) Neamati, N.; Mazumder, A.; Sunder, S.; Owen, J. M.; Tandon, M.; Lown, J. W.; Pommier, Y. *Mol. Pharmacol.* **1998**, *54*, 280-290.

- (22) Ryabinin, V. A.; Sinyakov, A. N.; de Soultrait, V. R.; Caumont, A.; Parissi, V.; Zakharova, O. D.; Vasyutina, E. L.; Yurchenko, E.; Bayandin, R.; Litvak, S.; Tarrago-Litvak, L.; Nevinsky, G. A. *Eur. J. Med. Chem.* **2000**, *35*, 989-1000.
- (23) Pommier, Y.; Neamati, N. *Adv. Virus Res.* **1999**, *52*, 427-458.
- (24) Pommier, Y.; Marchand, C.; Neamati, N. *Antiviral Res.* **2000**, *47*, 139-148.
- (25) Carteau, S.; Mouscadet, J. F.; Goulaouic, H.; Subra, F.; Auclair, C. *Biochem. Pharmacol.* **1994**, *47*, 1821-1826.
- (26) Wang, T.; Balakrishnan, M.; Jonsson, C. B. *Biochemistry* **1999**, *38*, 3624-3632.
- (27) Craigie, R. *J. Biol. Chem.* **2001**, *276*, 23213-23216.
- (28) Hyde, C. C.; Bushman, F. D.; Mueser, T. C.; Yang, Z.-N. *J. Mol. Biol.* **1999**, *296*, 535-538.
- (29) Wang, J.-Y.; Ling, H.; Yang, W.; Craigie, R. *EMBO J.* **2001**, *20*, 7333-7343.
- (30) Gao, K.; Butler, S. L.; Bushman, F. *EMBO J.* **2001**, *20*, 3585-3576.
- (31) Dervan, P. B.; Burli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688-693.
- (32) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.
- (33) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895.
- (34) Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A. *J. Biol. Chem.* **2000**, *275*, 24246-24254.
- (35) Gottesfeld, J. M., Turner, J. M., and Dervan, P. B. (2000) *Gene Expr* *9*, 77-91.
- (36) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
- (37) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.
- (38) Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 9354-9360.
- (39) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561.



- (40) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.
- (41) Felkner, R. H.; Roth, M. J. *J. Virol.* **1992**, *66*, 4258-4264.
- (42) Jonsson, C. B.; Donzella, G. A.; Roth, M. J. *J. Biol. Chem.* **1993**, *268*, 1462-1469.
- (43) LaFemina, R. L.; Callahan, P. L.; Cordingley, M. G. *J. Virol.* **1991**, *65*, 5624-5630.
- (44) Vink, C.; van Gent, D. C.; Elgersma, Y.; Plasterk, R. H. *J. Virol.* **1991**, *65*, 4636-4644.
- (45) Balakrishnan, M.; Jonsson, C. B. *J. Virol.* **1997**, *71*, 1025-1035.
- (46) Esposito, D.; Craigie, R. *EMBO J.* **1998**, *17*, 5832-5843.
- (47) Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J. *Biochemistry* **1999**, *38*, 2143-2151.
- (48) Bushman, F. D.; Craigie, R. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 1339-1343.
- (49) Pruss, D.; Reeves, R.; Bushman, F. D.; Wolffe, A. P. *J. Biol. Chem.* **1994**, *269*, 25031-25041.
- (50) Pruss, D.; Bushman, F. D.; Wolffe, A. P. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 5913-5917.
- (51) Pryciak, P. M.; Sil, A.; Varmus, H. E. *EMBO J.* **1992**, *11*, 291-303.
- (52) Pryciak, P. M.; Varmus, H. E. *Cell* **1992**, *69*, 769-780.
- (53) Pryciak, P. M.; Muller, H. P.; Varmus, H. E. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 9237-9241.
- (54) Miller, M. D.; Hazuda, D. *Curr. Opin. Microbiol.* **2001**, *4*, 535-539.
- (55) Carteau, S.; Mouscadet, J. F.; Goulaouic, H.; Subra, F.; Auclair, C. *Biochem. Biophys. Res. Commun.* **1993**, *192*, 1409-1414.
- (56) Fesen, M. R.; Kohn, K. W.; Leteurtre, F.; Pommier, Y. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2399-2403.
- (57) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.

- (58) Katzman, M., Katz, R. A. *Adv. Virus Res.* **1999**, 52, 371-395.
- (59) Davies, D. R.; Goryshin, I. Y.; Reznikoff, W. S.; Rayment, I. *Science* **2000**, 289, 77-85.
- (60) Roe, T.; Chow, S. A.; Brown, P. O. *J. Virol.* **1997**, 71, 1334-1340.
- (61) Jonsson, C. B.; Donzella, G. A.; Gaucan, E.; Smith, C. M.; Roth, M. J. *J. Virol.* **1996**, 70, 4585-4597.
- (62) Jonsson, C. B.; Roth, M. J. *J. Virol.* **1993**, 67, 5562-5571.
- (63) Scottoline, B. P.; Chow, S.; Ellison, V.; Brown, P. O. *Genes Dev.* **1997**, 11, 371-382.
- (64) Katz, R. A.; DiCandoloro, P.; Kukulj, G.; Skalka, A. M. *J. Biol. Chem.* **2001**, 276, 34213-34220.
- (65) deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. *E. J. Am. Chem. Soc.* **1997**, 119, 7909-7916.
- (66) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, 5, 104-109.
- (67) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, 295, 557-567.
- (68) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, 282, 111-115.
- (69) Chang, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, 122, 4856-4864.
- (70) Wurtz, N. R.; Dervan, P. B. *Chem. Biol.* **2000**, 7, 153-161.
- (71) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. *Proc. Natl. Acad. Sci. USA* **2000**, 97, 9367-9372.

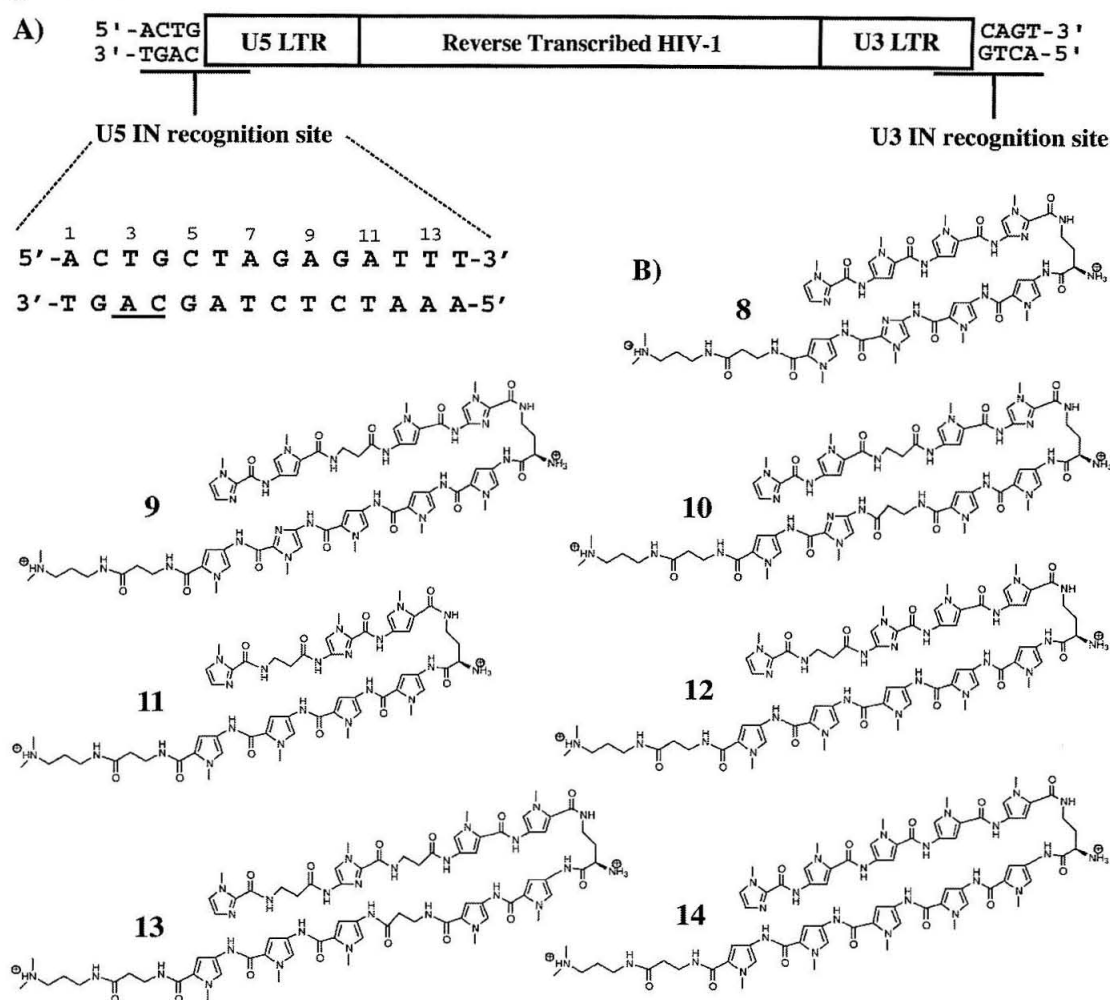
## Chapter 2B:

### Inhibition of HIV-1 Integration by Hairpin Polyamides <sup>1</sup>

#### Introduction

As discussed in Chapter 2A, hairpin polyamides are effective inhibitors of the *in vitro* integration of M-MuLV, a retrovirus related to HIV-1. HIV-1 integrase is increasingly the focus of antiviral research.<sup>2-11</sup> In addition to being an important therapeutic target, inhibition of HIV-1 integration is a promising goal for hairpin polyamides, given the positive cellular uptake and nuclear localization properties of polyamide-fluorescent dye conjugates in both cultured and primary human T-cells as compared to other cell-lines (Chapter 4). Previous studies have shown that polyamides targeted to HIV-1 promoter inhibited viral replication in human peripheral blood lymphocytes.<sup>12</sup> Thus, the current generation of hairpin polyamides, if they are shown to be effective HIV-1 integrase inhibitors, may have sufficiently favorable uptake and localization properties to be useful biochemical tools for understanding HIV-1 integration in cell culture, and as lead compounds for therapeutics. This chapter describes preliminary studies aimed at generating potent polyamide inhibitors of HIV-1 integration.

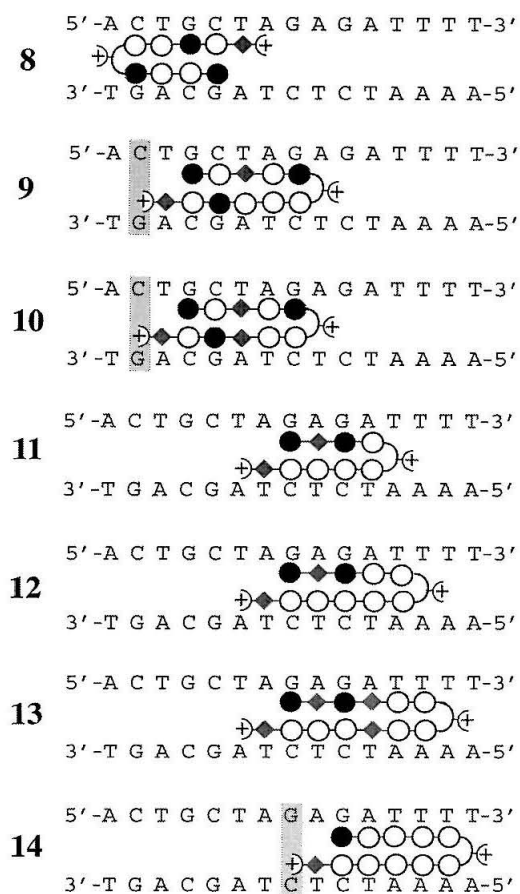
Unlike M-MuLV, in HIV-1 the two ends of the reverse-transcribed DNA are not the same, thus HIV-1 integrase has two different recognition sites, the termini of the U3 and U5 LTRs. Our studies focused on the terminal base pairs of the U5 LTR, 5'-ACTGCTAGAGATTTT-3' (Figure 2.9A). Seven polyamides (**8-14**) were designed to bind sites ranging from positions 1 to 15 of the HIV-1 U5 LTR (Figures 2.9B and 2.10). The affinities of polyamides for their target sites were measured by DNase I footprinting titrations using an internal copy of the HIV-1 U5 LTR terminus. Several of the compounds designed for the M-MuLV LTR were also investigated as mismatch controls. Polyamides were screened in preliminary cell-free assays for inhibition of 3' processing and strand transfer.

**Figure 2.9**

**Figure 2.9.** **A)** Schematic representation of the HIV-1 genome following reverse transcription. HIV-1 integrase recognizes two distinct sequences at the termini of the viral DNA. This work focuses on polyamides targeted to the U5 LTR terminus, sequence and numbering as shown. The conserved 5'-CA-3' residues are underlined. The two terminal bases to the 3' side of the CA dinucleotide, 5'-GT-3', are excised in the 3'-processing step by HIV-1 IN. **B)** Chemical structures of the polyamides targeted to the U5 LTR terminus.

## Results and Discussion

**Polyamide Design.** A general discussion of the pairing rules and other aspects of polyamide design is provided in Chapter 2A. Compared to the M-MuLV LTR terminus, the HIV-1 U5 LTR terminus has a greater G,C content on the terminal half of the sequence, and a greater A,T content on the internal half (compare Figure 2.10 to 2.1B).

**Figure 2.10**

**Figure 2.10.** Ball-and-stick representations of the polyamides and their binding sites within the HIV-1 U5 LTR terminus. Mismatches with the Dp tail are highlighted. For description of the ball and stick representation see Figure 2.1B.

Whereas in M-MuLV the internal half of the integrase recognition site was largely blocked from efficient recognition by the current generation of hairpin polyamides by the presence of five G•C base pairs in a row, this region was readily amenable to polyamide design in the HIV-1 U5 LTR (Figure 2.10). In particular, positions 6-12 were targeted with the known “Im-β-Im” polyamide **11**,<sup>13</sup> and positions 8-15 were targeted with the known 1-imidazole ten ring polyamide **14**.<sup>14</sup> Note that **14** is not a perfect match for this sequence as it places the Dp tail against a G•C base pair.<sup>15</sup> New polyamides **12** and **13** were designed as extensions of the Im-β-Im motif of **11** to larger binding sites. Polyamide **12** incorporates this motif into a ten-residue polyamide, targeting positions 6-13.

Polyamide **13** has an unusual Im-β-Im-β-PyPy N-terminal strand, paired with a 2-β-3 C-terminal strand, designed to target a nine base pair sequence (positions 6-14). Polyamides **8-10** were designed to occupy the center of the integrase recognition site, including the conserved 5'-CA-3' dinucleotide (positions 3-4). The β/Py pair containing polyamide **9** is the exact analog of polyamide **2** (Chapter 2A), both are targeted to positions 2-9 of their respective LTRs. Previously known polyamide **10** is the β/β pair analog of **9**.<sup>14</sup> However, unlike **2** with the M-MuLV sequence, **9** and **10** are not perfect match compounds, since the Dp tail is placed against a C•G base pair. Polyamide **8**,

which served as a mismatch compound for the M-MuLV study, was designed to bind at the absolute end of the U5 LTR, as well as the critical CA dinucleotide, covering positions 1-7.

**Quantitative DNase I Footprinting.** Quantitative DNase I footprinting titrations<sup>16</sup> were performed to determine the equilibrium association constants ( $K_a$ ) of polyamides **8-14** for their target sites in the HIV-1 U5 LTR, using a 194 bp DNA fragment derived from an HIV-1 circle junction (Table 2.3). Polyamides **2-3** were also investigated as mismatch controls. The circle junction was formed from the joining of the two ends of reverse-transcribed HIV-1 genome, such that the U5 and U3 LTR termini are adjacent internal sequences in the circle junction plasmid (HIV-cj).<sup>17</sup> This is a convenient plasmid for identifying integrase inhibitors, since both of the viral termini are present on the same fragment. Polyamide **8** bound its target site at the absolute end of the U5 LTR with reasonable affinity ( $K_a = 2 \times 10^9 \text{ M}^{-1}$ ) and excellent specificity (Figure 2.11). Polyamides **9** and **10** bound their match site with reasonably good affinity considering the Dp mismatch, with the 2- $\beta$ -2 polyamide **10** preferred by a factor of 4. However, the mismatch control polyamide **2** bound this sequence with only slightly lower affinity than **9**, and mismatch control polyamide **3** bound with the same affinity as **8** and **10**. Polyamide **2** is a single base pair mismatch of **9-10** as well as **14**, and binds both overlapping cognate sites (positions 2-15,  $K_a = 2 \times 10^8 \text{ M}^{-1}$ ). Polyamide **3** also bound this entire region. Polyamide **11** was specific for its match site, but showed low affinity equal to mismatch **2**. Polyamide **11** generally exhibits a  $\geq 10$ -fold higher  $K_a$ ,<sup>18</sup> the low affinity in this case is consistent with an energetic penalty for recognition of 5'-GA-3' steps as compared to 5'-GT-3' steps.<sup>19</sup> Polyamides **12** and **13** bind the same site as polyamide **11**, extended internally by one and two base pairs, respectively. Very unusual behavior was observed at this site for polyamide **12**, in that a hypersensitive cleavage band appeared to grow in within the footprint, precluding reliable quantitation at this site.

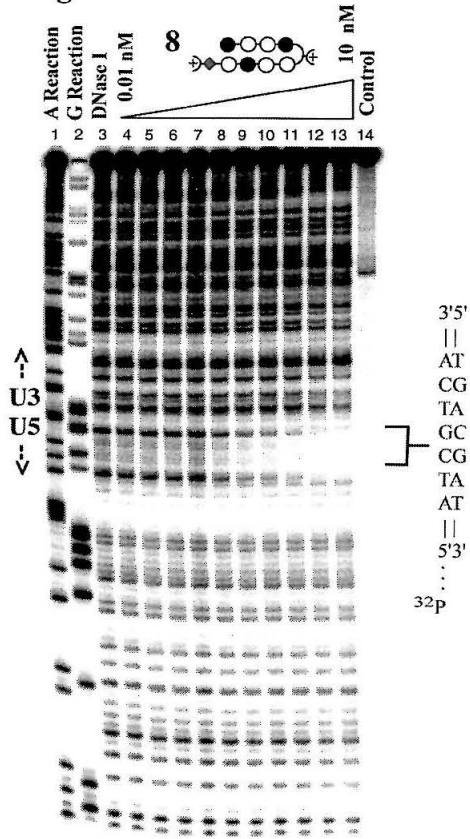
**Table 2.3.** Equilibrium association constants for an internal copy of the HIV-1 U5 LTR terminus.

Polyamide	Sequence	K <sub>a</sub> (M <sup>-1</sup> )
8	5'-agtACTGCTAGag-3'	2 x 10 <sup>9</sup>
9	5'-gtaCTGCTAGAgag-3'	5 x 10 <sup>8</sup>
10	5'-gtaCTGCTAGAgag-3'	2 x 10 <sup>9</sup>
11	5'-tgcTAGAGATttt-3'	1 x 10 <sup>8</sup>
12	5'-tgcTAGAGATTttc-3'	~1 x 10 <sup>10</sup> <sup>a</sup>
13	5'-tgcTAGAGATTTtcc-3'	2 x 10 <sup>10</sup>
14	5'-ctaGAGATTTTcca-3'	1 x 10 <sup>10</sup>
2	5'-gtaCTGCTAGAgag-3' <sup>b</sup> 5'-ctaGAGATTTTcca-3'	2 x 10 <sup>8</sup>
3	5'-ccaGTACTGCTAGAgag-3' <sup>b</sup> 5'-ctgCTAGAGATTTTcca-3'	2 x 10 <sup>9</sup>

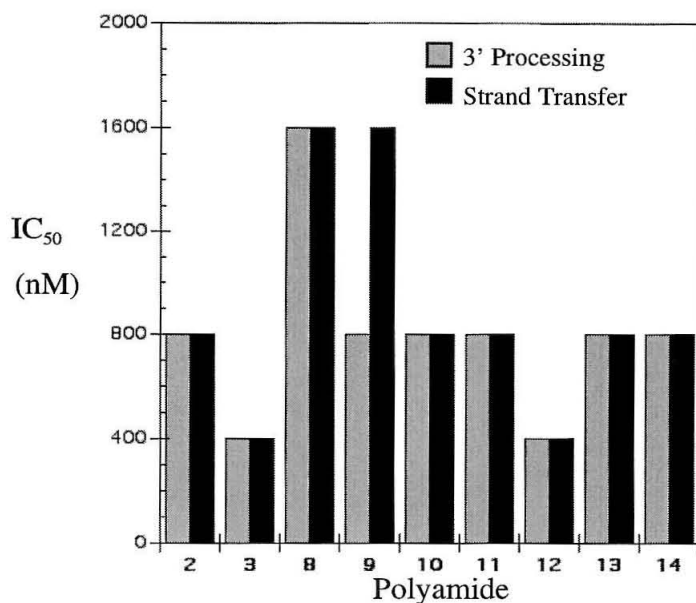
**Table 2.3.** Values reported are the mean values obtained from at least two DNase I footprint titration experiments. Mismatched residues are underlined. The assays were carried out at 22° C, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. (a) This value is an estimate, not quantitative (see text). (b) Overlapping sites.

**Figure 2.11.** Quantitative DNase I footprint titration experiment with polyamide **8** on a 5' <sup>32</sup>P labeled 194 base pair DNA fragment containing an internal copy of the HIV-1 U5 LTR terminus: lane 1, A reaction; lane 2, G reaction; lane 3, DNase I standard; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide, respectively; lane 14, intact DNA. All reactions contained 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>, and were performed at 22° C.

**Figure 2.11**



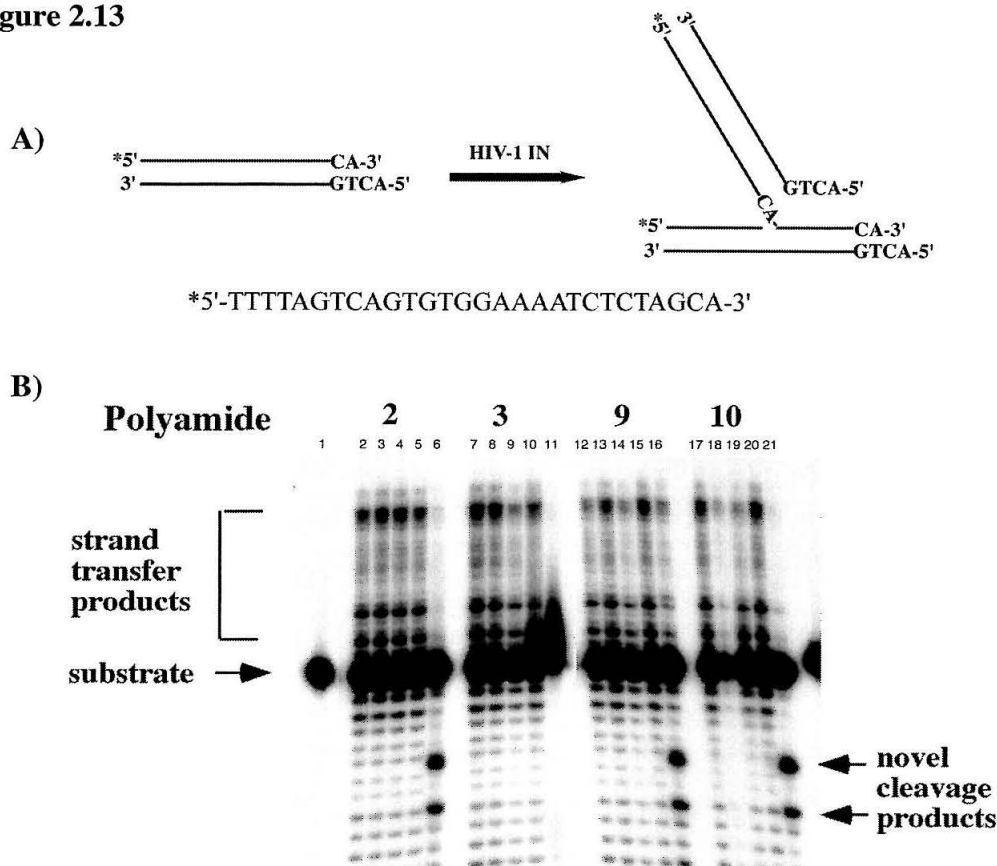
Visual inspection of the gels suggests that the affinity ( $K_a \sim 1 \times 10^{10} \text{ M}^{-1}$ ) is considerably higher than for polyamide **11**. The large polyamide **13** bound this region with high affinity without exhibiting unusual behavior, however, it coated the DNA (bound to all sequences uniformly) at relatively low concentrations. Polyamide **14** bound its match site with high affinity ( $K_a = 1 \times 10^{10} \text{ M}^{-1}$ ) and good specificity.

**Figure 2.12**

**Figure 2.12.** Preliminary IC<sub>50</sub> of polyamides on oligonucleotide-based HIV-1 integrase activities. Each polyamide (25—1600 nM) was titrated in either the 3' processing (grey) or strand transfer (black) assays. IC<sub>50</sub> values were determined as the concentration of each polyamide that inhibited IN activities by 50%.

***In Vitro* HIV-1 Integration Assays.** Preliminary cell free assays of HIV-1 integrase activity were performed in analogy to the assays reported for M-MuLV integration in Chapter 2A. It should be noted that the M-MuLV IN assays were following overnight incubation of the polyamides with the substrate DNA. The preliminary HIV-1 IN assays were performed with a much shorter pre-incubation time ( $\leq 1$  hour). The IC<sub>50</sub> values for 3' processing and strand transfer ranged from a moderate 400 nM to  $\geq 1600$  nM (Figures 2.12 and 2.13). A major conclusion of Chapter 2A is that polyamides which bind at the termini of the M-MuLV LTR were much more effective inhibitors of the 3' processing reaction than the strand transfer reaction, while polyamides that bound internally through position 9 had similar efficacy for both reactions. Polyamide **8** was the only polyamide designed for the absolute end of the HIV-1 U5 LTR, and unfortunately was a poor inhibitor of both 3' processing and strand transfer



**Figure 2.13**

**Figure 2.13.** A) Schematic of the strand transfer assay with the sequence of the labeled oligonucleotide shown below. HIV-1 IN is added to the precleaved duplex which yields  $^{32}\text{P}$ -labeled products larger than the input substrate, as well as smaller products from the invaded strand. B) An HIV-1 IN strand transfer assay with polyamides **2,3,9,10**, highlighting the enhanced bands resulting from novel cleavage of the substrate or altered target site selection. Lane 1, no protein control; lanes 2,7,12, and 17 HIV-1 IN alone. Lanes 3-6, IN plus polyamide **2**; lanes 8-11, IN plus polyamide **3**; lanes 13-16, IN plus polyamide **9**; lanes 18-21, IN plus polyamide **10**. The concentrations of the polyamide used are (left to right): 1, 10, 100 and 1000 nM.

( $\text{IC}_{50} \geq 1600$  nM). Polyamide **9** was the only polyamide tested which displayed a difference in inhibition between assays, favoring the 3'-processing assay by a factor of 2. The majority of polyamides tested were 800 nM inhibitors in both assays. Mismatch control **2** had the same values as match compounds **9-11** and **13-14** in the 3' processing assay, and outperformed its analog **9** in strand transfer inhibition. Similarly, extended hairpin **3**, also intended as a mismatch control, was one of only two 400 nM inhibitors.

The other moderately potent inhibitor was polyamide **12**, which displayed both high affinity and unusual behavior in the DNase I footprinting experiments.

Figure 2.13B shows a set of strand transfer assays highlighting novel cleavage bands which were observed for some, but not for all of the polyamides tested. These bands may be the result of direct cleavage of the substrate DNA by the nuclease activity of HIV-1 at a site other than the 5'-CA-3' dinucleotide, or they may represent altered target site selection events, as seen for M-MuLV. However, further work is necessary to define this. Note that for **2** and **10**, the novel cleavage bands occurred at the same (high) concentration as inhibition of the normal strand transfer products, but for **9** inhibition did not occur at an appreciable extent, even though the novel cleavage bands are quite intense. Independent of inhibition, the ability of polyamides to alter target site selection or to induce direct cleavage of the substrate viral DNA at unnatural sites suggests exciting new roles for polyamides in HIV-1 integration research.

Overall the preliminary results are encouraging, although similar compounds appear to be somewhat less potent for HIV-1 integration inhibition than M-MuLV integration inhibition. This may reflect structural differences in the integrases themselves,<sup>20</sup> differences in substrate recognition by the integrases,<sup>21-23</sup> potential differences in oligomeric state of the proteins on and off the DNA,<sup>24-28</sup> or, simply that optimal polyamides have not yet been generated for the HIV-1 U5 LTR terminus. Lower inhibition for HIV-1 than M-MuLV with polyamides targeted to the same positions with essentially the same affinity (compare **8**(HIV-1) with **7**(M-MuLV), and **10**(HIV-1) with **1**(M-MuLV)), is consistent with studies of modified LTRs containing nucleotide analogs that suggested that changes to the minor groove had a lesser affect on HIV-1 integrase activity than M-MuLV integrase activity.<sup>22</sup> Thus, it may require polyamides with higher affinities to reach the same levels of inhibition.

In terms of optimizing binding affinity, polyamides **9**, **10** and **14** might benefit from replacement of the  $\beta$ -Dp tail by a shorter derivative such as methylamide (see

Chapter 5), which would eliminate Dp mismatches. However, in terms of inhibition, **10** would probably benefit from an extended tail rather than a truncated tail, since **3** (which has the C-terminal extension), even as a mismatch was one of the two most potent inhibitors in this study, and for M-MuLV it was most potent in the more demanding assay of strand transfer into an exogenous target. The M-MuLV study also suggests that the ten-ring analog of **9** and **10**, potentially with the  $\gamma$ -turn instead of the  $(R)^{H_2N}\gamma$ -turn, may be an especially effective inhibitor (in analogy to **1**).

A hypothesis from the M-MuLV study was that polyamides targeting internal sites such as **14** might be better strand transfer inhibitors than 3'-processing inhibitors. While this does not appear to be the case for HIV-1 integrase, the fact that a polyamide which binds so far away from the viral terminus is an inhibitor at all suggests that good inhibition may be obtained by targeting this internal site and the viral terminus simultaneously with two polyamides binding nonoverlapping sites. Toward this end, a methylamide tail for **14** would be especially beneficial as it would open up an extra base pair in the target region for a second polyamide to bind. It would also be interesting to investigate truncated tail analogs of the most potent match inhibitor **12**, which would open up a base pair in the target region and may shed light on the usual behavior of **12** with regards to DNase I footprinting.

In order to target two polyamides to the HIV-1 U5 LTR, better options than **8** will have to be found for the LTR terminus. Here, the affinity for the substrate DNA will be lower than for the  $K_a$  obtained by footprinting analysis (see Chapter 2A appendix), and only compounds (**5,6**) with equilibrium association constants above  $10^{10} \text{ M}^{-1}$  were truly effective inhibitors of M-MuLV integration. It was noted these compounds were not effective strand transfer inhibitors for M-MuLV, and it will be interesting to see if tighter binders for the terminal portion of the HIV-1 U5 LTR behave similarly.

## Conclusions

The polyamides described here, while at present are only moderate inhibitors of the *in vitro* reactions of HIV-1 integrase, are valuable lead compounds for further development. Particularly noteworthy was the high activity of mismatch polyamide **3**, which argues that a C-terminal extension at the viral terminus is beneficial for inhibition, and the inhibitory activity of polyamides which bind internally from the viral terminus such as **12** and **14**. With the development of higher affinity compounds targeted to the terminus, this should allow for synergistic inhibition by two polyamides, which has been successful in other studies on HIV-1 inhibition.<sup>12</sup> Thus far, the U3 LTR terminus remains an unexplored target for polyamide inhibition of HIV-1 integration. In addition to polyamides with improved binding affinity and placement within the U5 and U3 LTRs, potent inhibitors may be obtained by modified polyamides which could be designed to deliver known integrase inhibitors to the enzyme, or to trap the terminal nucleotides by alkylation, as discussed in Chapter 2A. Additionally, “positive patch” polyamides<sup>29-31</sup> may be able to directly target the phosphate group that is attacked by HIV-1 integrase in the 3' processing step. Finally, the most interesting result of the current study is the novel cleavage bands that are observed in the *in vitro* assays with some, but not all, of the polyamides, and which did not appear to be directly related to inhibition. Further study of this unusual phenomenon may outline new roles for polyamides in the study of HIV-1 integration.

## Experimental

**Polyamides.** All polyamides were synthesized by solid-phase methods as previously described,<sup>32,33</sup> and characterized by analytical HPLC and MALDI-TOF mass spectrometry. For polyamides **1-8**, see Chapter 2A. ImPy-β-PyIm-(R)-<sup>H2N</sup>γ-PyPyPyIm-Py-β-Dp (**9**) [M+H] 1431.84, 1431.68 calculated for [M+H]. ImPy-β-PyIm-(R)-<sup>H2N</sup>γ-PyPy-β-ImPy-β-Dp (**10**) [M+H] 1380.83, 1380.67 calculated for [M+H]. Im-β-ImPy-

(R)-<sup>H2N</sup>γ-PyPyPyPy-β-Dp (**11**) [M+H] 1186.65, 1186.57 calculated for [M+H]. Im-β-ImPyPy-(R)-<sup>H2N</sup>γ-PyPyPyPyPy-β-Dp (**12**) [M+H] 1430.74, 1430.67 calculated for [M+H]. Im-β-Im-β-PyPy-(R)-<sup>H2N</sup>γ-PyPy-β-PyPyPy-β-Dp (**13**) [M+H] 1572.97, 1572.74 calculated for [M+H]. ImPyPyPyPy-(R)-<sup>H2N</sup>γ-PyPyPyPyPy-β-Dp (**14**) [M+H] 1480.82, 1480.60 calculated for [M+H].

**DNase I footprinting.** DNase I footprinting reactions were performed as previously described,<sup>16</sup> using a 194 base pair 5'-<sup>32</sup>P-labeled DNA fragment derived from an HIV-1 circle junction, such that the HIV-1 U5 and U3 LTR terminal sequences are connected in the center of the fragment. The labeled fragment was generated by the PCR method<sup>14,16</sup> using primers 5'-GTTGTGTGACTCTCGGTA ACTA and 5'-GTCAGTGGAT ATCTGATCCT and an HIV-1 circle junction-containing plasmid (HIV-cj, provided by the Roth group)<sup>17</sup> as the template. Primers were prepared and purified by the Caltech Biopolymer Synthesis Facility. The sequence of the labeled fragment is listed here with the terminal 14 base pairs of the HIV-1 U5 LTR shown in bold: 5'-<sup>32</sup>P GTTGTGTGACT CTCGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTA **GCAGT**ACTGGAAGGGCTAATTCAC TCCCAACGAAGACAAGATATCCTTGATC TGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACC AGGGCCAGGGATCAGATATCCACTGAC.

***In vitro* assays.** Cell-free integration assays were performed by Fan Yang in the Roth group using standard methods<sup>34</sup> analogous to the reactions described in Chapter 2A for M-MuLV integration.

## Acknowledgements

We thank the NIH for research support and are grateful to the Ralph M. Parsons Foundation for a predoctoral fellowship to J.M.B.

## References and Notes

- (1) Compound numbering follows from Chapter 2A.
- (2) See Chapter 2A and references therein. See also references 3-4 (reviews), 5-6 (new natural product inhibitors), 7-9 (new dinucleotide, oligonucleotide, and peptide inhibitors, respectively), and 10-11 (computation models).
- (3) Craigie, R. *J. Biol. Chem.* **2001**, 276, 23213-23216.
- (4) Chondra, J. H.; Miller, M. D.; Hazuda, D. J.; Emini, E. A. *Annu. Rev. Med.* **2002**, 53, 541-555.
- (5) Singh, S. B.; Zink, D. L.; Heimbach, B.; Genilloud, O.; Teran, A.; Silverman, K. C.; Lingham, R. B.; Felock, P.; Hazuda, D. J. *Org. Lett.* **2002**, 4, 1123-1126.
- (6) Singh, S. B.; Zink, D. L.; Bills, G. F.; Pelaez, F.; Teran, A.; Collado, J.; Silverman, K. C.; Lingham, R. B.; Felock, P.; Hazuda, D. J. *Tetrahedron Lett.* **2002**, 43, 1617-1620.
- (7) Taktakishvili, M.; Neamati, N.; Pommier, Y.; Pal, S.; Nair, V. *J. Am. Chem. Soc.* **2000**, 122, 5671-5677.
- (8) Brodin, P.; Pinskaya, M.; Buckle, M.; Parsch, U.; Romanova, E.; Engels, J.; Gottikh, M.; Mouscadet, J.-F. *Biochemistry* **2002**, 41, 1529-1538.
- (9) Maroun, R. G.; Gayet, S.; Benleulmi, M. S.; Porumb, H.; Zargarian, L.; Merad, H.; Leh, H.; Mouscadet, J.-F.; Troalen, F.; Fermandjian, S. *Biochemistry* **2001**, 40, 13840-13848.
- (10) Keseru, G. M.; Kolossvary, I. *J. Am. Chem. Soc.* **2001**, 123, 12708-12709.
- (11) Buolamwini, J. K.; Assefa, H. *J. Med. Chem.* **2002**, 45, 841-852.
- (12) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 12890-12895.
- (13) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Ghazal, P.; Dervan, P. B.; Gottesfeld, J. M. *Biochemistry* **1999**, 38, 10801-10807.

- (14) Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A. *J. Biol. Chem.* **2000**, *275*, 24246-24254.
- (15) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.
- (16) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.
- (17) Smith, J. S.; Kim, S.; Roth, M. J. *J. Virol.* **1990**, *64*, 6286-6290.
- (18) Dervan, P. B., Unpublished results and reference 13.
- (19) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532-12537.
- (20) Yang, F.; Seamon, J. A.; Roth, M. J. *Virology* **2001**, *291*, 32-45.
- (21) Morgan, A. L.; Katzman, M. J. *Gen. Virol.* **2000**, *81*, 839-849.
- (22) Wang, T.; Balakrishnan, M.; Jonsson, C. B. *Biochemistry* **1999**, *38*, 3624-3632.
- (23) Heuer, T. S.; Brown, P. O. *Biochemistry* **1997**, *36*, 10655-10665.
- (24) Gao, K.; Bulter, S. L.; Bushman, F. *EMBO J.* **2001**, *20*, 3565-3576.
- (25) Yang, F.; Roth, M. J. *J. Virol.* **2001**, *75*, 9561-9570.
- (26) Deprez, E.; Tuc, P.; Leh, H.; Mouscadet, J.-F.; Auclair, C.; Brochon, J.-C. *Biochemistry* **2000**, *39*, 9275-9284.
- (27) Yang, F.; Leon, O.; Greenfield, N. J.; Roth, M. J. *J. Virol.* **1999**, *73*, 1809-1817.
- (28) Heuer, T. S.; Brown, P. O. *Biochemistry* **1998**, *37*, 6667-6678.
- (29) Bremer, R.E.; Wurtz, N.R.; Szewczyk, J.W.; Dervan, P.B. *Bioorg. Med. Chem.* **2001**, *9*, 2093-2103.
- (30) Bremer, R.E.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1998**, *5*, 119-133.
- (31) Cashin, A.; Edelson, B.; Dervan, P.B. Unpublished results.
- (32) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
- (33) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.
- (34) Marchard, C.; Neamati, N.; Pommier, Y. *Methods Enzymol.* **2001**, *340*, 624-633.

## **CHAPTER 3**

### **DNA Binding Polyamides for the Inhibition of HER2 Transcription**

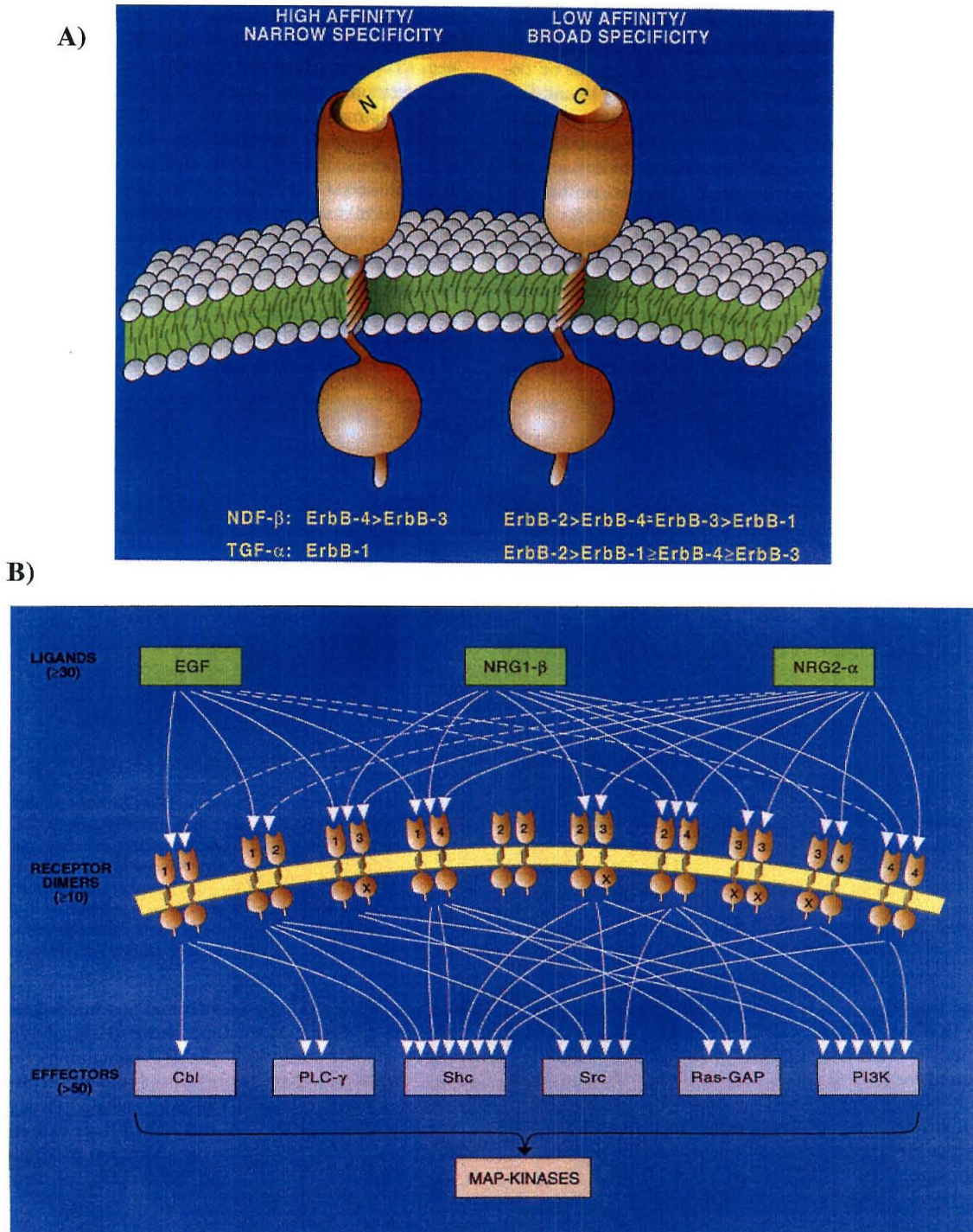


## Abstract

Overexpression of the growth factor receptor/tyrosine kinase HER2 is observed in approximately 30% of human breast cancers. HER2 overexpression is associated with tumor metastasis and resistance to chemotherapy. HER2 is itself an important target for chemotherapy, and has been the focus of antisense and antigene therapies. A comprehensive series of DNA binding polyamides has been designed to bind proximally to the TATA box on the HER2 promoter. The affinity and specificity of the polyamides for the target site has been determined and compared with related site. The same series of polyamides behaves quite differently in these two ostensibly similar contexts. For the HER2 site, high-affinity compounds were obtained which were strong inhibitors of TBP-binding and *in vitro* transcription of the HER2 gene. HER2 transcription inhibition in cell culture remains an ongoing goal.

## Introduction

The growth factor receptor/tyrosine kinase HER2 (also known as *neu* and erbB-2) is overexpressed in approximately 30% of human breast cancers.<sup>1,2</sup> HER2 overexpression is associated with tumor metastasis and resistance to chemotherapy.<sup>3</sup> In addition to breast cancer, HER2 has also been linked to ovarian and lung cancers, wherein overexpression correlates with shorter patient survival time.<sup>4</sup> Overexpression does not appear to result from a transforming mutation in the HER2 gene itself, suggesting that the wild type is oncogenic under these conditions. Overexpression has been shown to result from both gene amplification<sup>5</sup> and transcriptional upregulation.<sup>6</sup> HER2 is a member of the epidermal growth factor (EGF) class of receptor/tyrosine kinases, which includes epidermal growth factor receptor (EGFR, erbB-1), erbB-3, and erbB-4.<sup>4,7</sup> These four receptors are known to interact with over thirty growth factor ligands such as EGF and transforming growth factor (TGF)- $\alpha$ . Two populations of receptors for the same growth factor are typically observed: a low-affinity population, which has been interpreted as the monomer, and a high-affinity population, which has been interpreted as the dimer, although this generally accepted view has been recently challenged.<sup>8,9</sup> A HER2-specific ligand has not yet been isolated, and indeed such a ligand may not exist. Instead, HER2 appears to be a shared, low-affinity receptor (co-receptor) for multiple growth factors which HER2 binds as a heterodimer with each of the other three erbB receptors (Figure 3.1).<sup>4</sup> HER2 is the preferred heterodimerization partner among the erbB receptors and acts as a potent tyrosine kinase in the heterodimer.<sup>7</sup> The erbB receptors also form homodimers which are active in signal transduction, although the physiological role of the HER2 homodimer is a subject of debate.<sup>10</sup> With input from more than thirty growth factor ligands, the four erbB receptors activate more than fifty effectors.<sup>4</sup> All of these effectors appear to activate MAP-kinases, which ultimately trigger cellular growth and differentiation. It is readily apparent that disturbing this complex regulatory network by overexpressing the most permissive of the erbBs can lead to cancer. Among the down-

**Figure 3.1**

**Figure 3.1.** HER2 (also known as *neu*, erbB-2) is an EGF receptor/tyrosine kinase, which acts as a low-affinity co-receptor for ligands such as TGF-α, and as a co-activator for effectors such as Ras-GAP, which ultimately activate MAP-kinases, leading to cell growth and differentiation. **A)** Schematic of ligand recognition by erbB dimers. **B)** Overview of the erbB signaling network.<sup>4</sup>

stream targets of HER2 are promoters of metastasis and angiogenesis, which are likely responsible for the aggressively malignant phenotype observed in HER2-overexpressing breast cancers.<sup>6,11-13</sup> Additionally, several of the downstream targets are transcription factors which upregulate transcription of the HER2 gene itself, largely through the intermediacy of the Ets transcription factor ESX, resulting in positive feedback.<sup>13-15</sup> Thus, even an initially minimal amount of HER2 overexpression can have considerable transforming potential.

HER2 is an important target for breast cancer chemotherapy. In particular, trastuzumab, a humanized monoclonal antibody targeting the HER2 receptor, has proven to be highly effective for the treatment HER2-positive breast cancer, particularly in cases with high overexpression of the HER2 gene, both as a first-line therapy and in combination with more traditional chemotherapeutic agents.<sup>16,17</sup> Trastuzumab (marketed as Herceptin), one of the first monoclonal antibodies used clinically for human cancer therapy, is widely considered a biotechnology success story. The clinical efficacy of trastuzumab has prompted considerable interest in additional HER2-related therapeutics, including second generation antibodies<sup>18,19</sup> and peptides derived from the HER2 receptor itself as tumor vaccines.<sup>20</sup>

There is also significant interest in HER2 as a target of antigene and antisense studies.<sup>21</sup> The typical HER2-overexpressing, aggressively malignant tumor has two to ten copies of the HER2 gene and one million copies of the HER2 receptor, providing a mathematical rationale for antisense or antigene therapy. Antisense oligonucleotides have been shown to reduce HER2 mRNA and protein levels in a dose-dependent, sequence-specific manner.<sup>22,23</sup> Oligonucleotide-induced down-regulation resulted in inhibition of cell-cycle progression at G<sub>1</sub> and apoptotic cell death of cultured HER2-overexpressing breast cancer cells. Induction of apoptosis was synergistically enhanced by the co-administration of conventional chemotherapeutics (which these HER2-over-

**Figure 3.2****The HER2 Promoter**

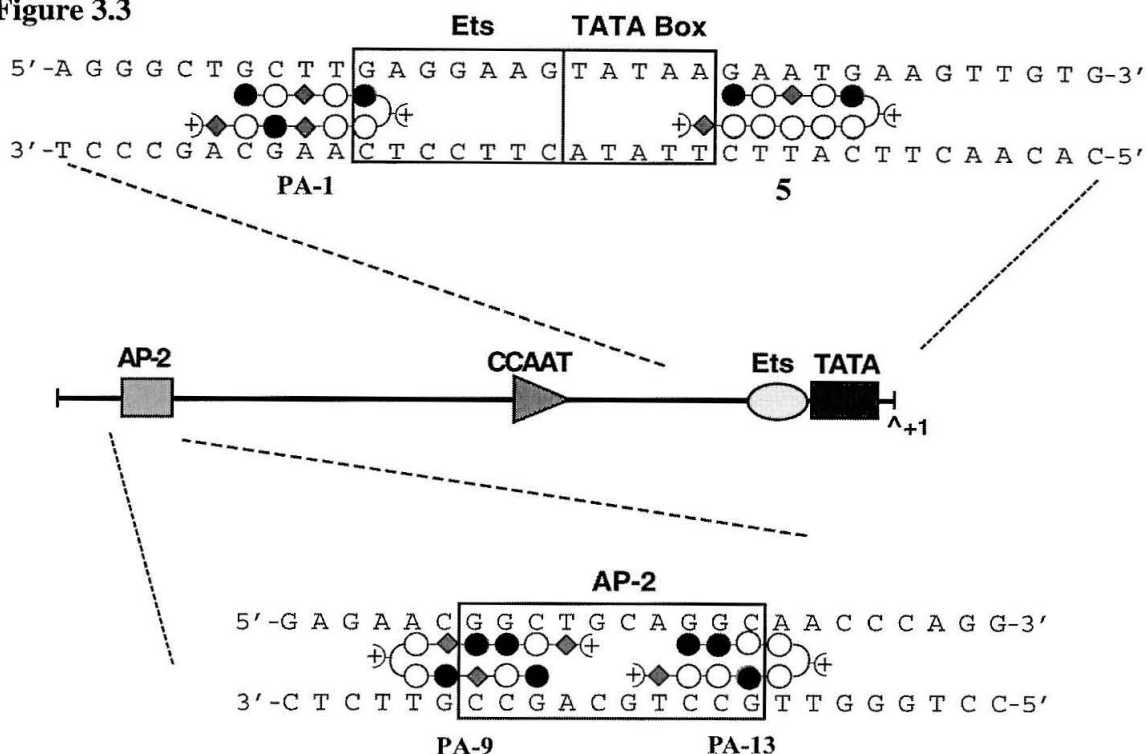
5' - CCCGGGGGTCCTGGAAGCCACAAGGTAAACACAACACATCCCCCTCCTTGAATATGCAATTTTACTA  
 GAGGATGTGGTGGGAAAACCATTTATTTGATATTAAAACAAATAGGCTTGGGATGGAGTAGGATGCAAGCT  
 CCCCAGGAAAGTTTAAGATAAAACCTGAGACTTAAAAGGGTGTTAAGAGTGCCAGCCTAGGGAATTTATC  
 CCGGACTCCGGGGGAGGGGGCAGAGTCACCAGCCTCTGCATTTAGGGATTCTCCGAGGAAAAGTGTGAGA  
**AP-2**  
 ACGGCTGCAGGCAACCCAGGCGTCCCGGCGCTAGGAGGGACGACCCAGGCCTGCGCGAAGAGAGGGAGAA  
 AGTGAAGCTGGGAGTTGCCGACTCCCAGACTTCGTTGGAATGCAGTTGGAGGGGGCGAGCTGGGAGCGCG  
**CCAAT box Polypurine/polypyrimidine mirror repeat Ets TATA box**  
 CTTGCTCCCAATCACAGGAGAAGGAGGAGGTGGAGGAGGAGGGCTGCTTGAGGAAGTATAAGAATGAAGT  
**ZONAB**  
 TGTGAAGCTGAGATTCCCTCCATTGGGACCGGAGAAACAGGGGAGCCCCCGGGCAGCCGCGCGCCCC  
<sup>^</sup><sub>+1</sub>  
 TTCCACGGGGCCCTTTACTGCGCCGCGCGCCCGCCCCACC - 3'

**Figure 3.2.** Sequence of the HER2 promoter<sup>24</sup> with the main transcription factor binding sites (positive regulators AP-2, CCAAT box, Ets, and the TATA box, and repressor ZONAB), structural elements (the polypurine/polypyrimidine mirror repeat), and the major transcription start site (+1) shown.<sup>21</sup> Translation of the protein begins at +178 relative to the major transcription start site.

expressing cells are resistant to in the absence of antisense oligonucleotides). Triplex-forming oligonucleotides have been used to successfully target a polypurine/polypyrimidine minor repeat which is an important structural regulatory element in the HER2 promoter<sup>24</sup> (Figure 3.2), resulting in transcription inhibition in cells when administered with lipofection.<sup>25</sup> The HER2 promoter has also been successfully targeted by designed zinc finger transcription factors.<sup>26</sup> In a powerful demonstration of designed gene regulation, fusions of the designed zinc finger with transcriptional repressor and activator domains were able to repress and activate, respectively, HER2 transcription in cell-culture.<sup>27</sup> Repression in HER2-overexpressing breast cancer cells by the designed zinc finger transcription factors resulted in G<sub>1</sub> accumulation, similar to the inhibition of cell-cycle progression observed with antisense oligonucleotides. Together these studies demonstrate the potential of transcriptional and posttranscriptional regulation for the treatment of HER2-overexpressing human breast cancers.



The Dervan laboratory has also been interested in the HER2 promoter, as part of a program aimed at the development of small molecule modulators of gene regulation.<sup>28,29</sup> Polyamides composed of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamides can be designed to bind specifically to a large number of predetermined DNA sequences, with affinities that rival natural transcription factors.<sup>30,31</sup> In a hairpin polyamides, the N- and C-terminal strands of Py/Im aromatic amino acids are connected by an alkyl amino acid—either  $\gamma$ -aminobutyric acid ( $\gamma$ ), or the chiral, amine-functionalized derivative (*R*)-2,4,-diaminobutyric acid ( $((R)^{H_2N}\gamma)$ )—and bind side-by-side in the minor groove of DNA with an antiparallel orientation relative to each other and a 5'-3' N-C orientation relative to DNA. This binding motif results in amino acid pairings in the minor groove that stack the aromatic Py/Im rings against each other and the walls of the groove allowing backbone amide hydrogens and the substituents at the 3-position of the polyamide residues to make specific contacts with the edges of the DNA base pairs. These specific contacts translate into a set of pairing rules for DNA recognition by hairpin polyamides. A Py/Im pair targets a C•G base pair, an Im/Py pair targets a G•C base pair, and a Py/Py pair binds to an A•T or T•A base pair. The *N*-methyl-3-hydroxypyrrole (Hp) residue has been developed to break the degeneracy of the Py/Py pair. Hairpin polyamides which bind six to seven base pair sites with high affinity ( $10^9$  -  $10^{10}$  M<sup>-1</sup>) have been shown to be effective in cell-free assays at the inhibition of transcription by interfering with transcription factor binding.<sup>29</sup> Polyamides have also been shown to bind their cognate sites on isolated nucleosomes.<sup>32</sup> *In vivo*, polyamides targeted to transcription factor binding sites in the HIV-1 promoter resulted in inhibition of retroviral replication in human cells.<sup>33</sup> Two polyamides targeted to sequences proximal to the DNA recognition sites for the ETS-1, LEF-1, and TBP transcription factors in the HIV-1 promoter acted in synergy at 1  $\mu$ M each to obtain >99.9% inhibition of retroviral replication.

**Figure 3.3**

**Figure 3.3.** Examples of polyamides targeted to transcription factor binding sites on the HER2 promoter. The promoter is shown schematically in the center beginning with the major transcription start site (+1) on the right, with the positions of transcription factor binding sites shown to scale. The sequences shown above and below highlight the transcription factor binding sites (boxes) and targeted polyamides (ball-and-stick models). The symbols for polyamides are open, filled, and slashed circles, Im, Py and Bz rings, respectively; curved lines,  $\gamma$ -aminobutyric acid; diamonds,  $\beta$ -alanine; dimethylaminopropylamine tail and amino group of (*R*)-2,4-diaminobutyric acid, plus signs. Polyamides designated PA (**PA-1**, **PA-9**, **PA-13**) were prepared by Roland Burli. Polyamide **5**, this work.

In addition to the therapeutic potential of targeting the HER2 promoter, inhibition of HER2 transcription *in vivo* would demonstrate the efficacy of polyamides for regulation of an endogenous human gene in human cells. Polyamides have been targeted to three of the most important transcription factor binding sites in the HER2 promoter, the AP-2,<sup>34</sup> Ets,<sup>28</sup> and TBP binding sites (Figures 3.2 and 3.3). This work focuses on the TATA box, the recognition element for the TATA Binding Protein (TBP), a saddle-shaped protein that binds in the minor groove of DNA, causing a large bend in the double helix (Figure 3.4).<sup>35</sup> TBP is the DNA binding subunit of the TFIID transcription factor, which nucleates the assembly of the RNA polymerase II transcription machinery. The

**Figure 3.4**

**Figure 3.4.** The TATA-binding protein (TBP, TFIID). TBP2 from *Arabidopsis thaliana* shown complexed with the *Adenovirus* major late promoter TATA element.<sup>35</sup> +1 represents the transcription start site.

TATA box is a ubiquitous transcriptional element found in many human promoters.<sup>36</sup> Minor-groove contacts are made on both sides of the consensus TATA sequence, and polyamides targeted to proximal sequences are likely to inhibit TBP binding by direct competition for the minor groove, and by allosteric stabilization of B-form DNA as compared to the TBP-induced bend.<sup>29</sup> Polyamides targeted to sequences proximal to the TATA box of a given promoter may be promoter-specific inhibitors of transcription, rather than genome-wide inhibitors of TBP binding. This approach was previously used for the HIV-1 promoter, where polyamides targeting sequences proximal to the TATA box inhibited TBP binding and *in vitro* transcription as well as viral replication *in vivo*.<sup>33</sup> The mRNA levels of seven other genes with different proximal sequences to their TATA boxes were not effected by the polyamide treatment.

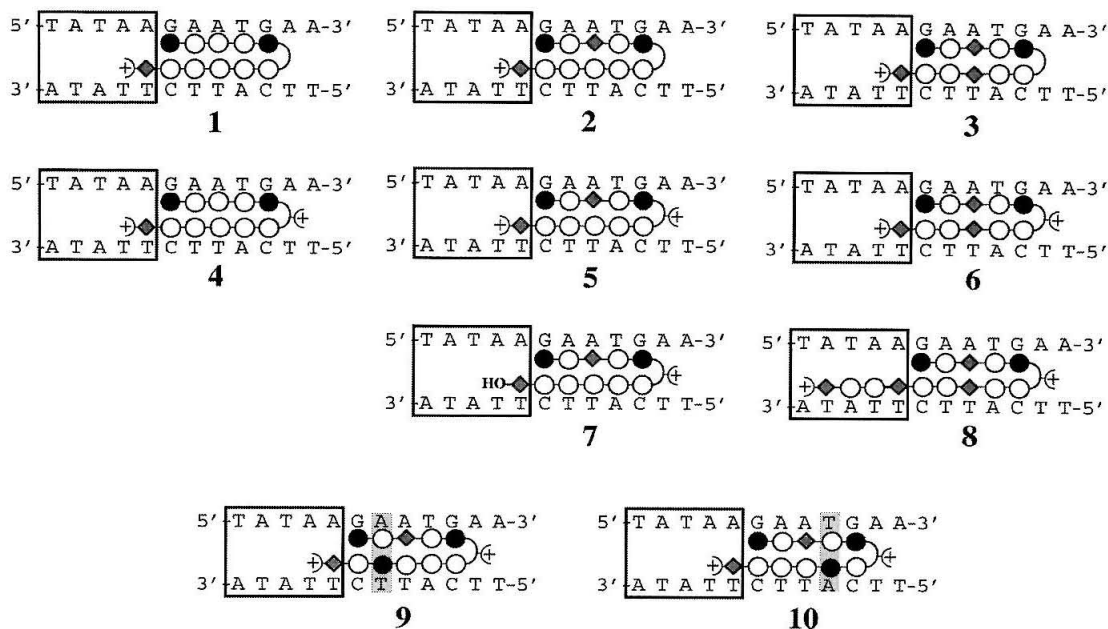
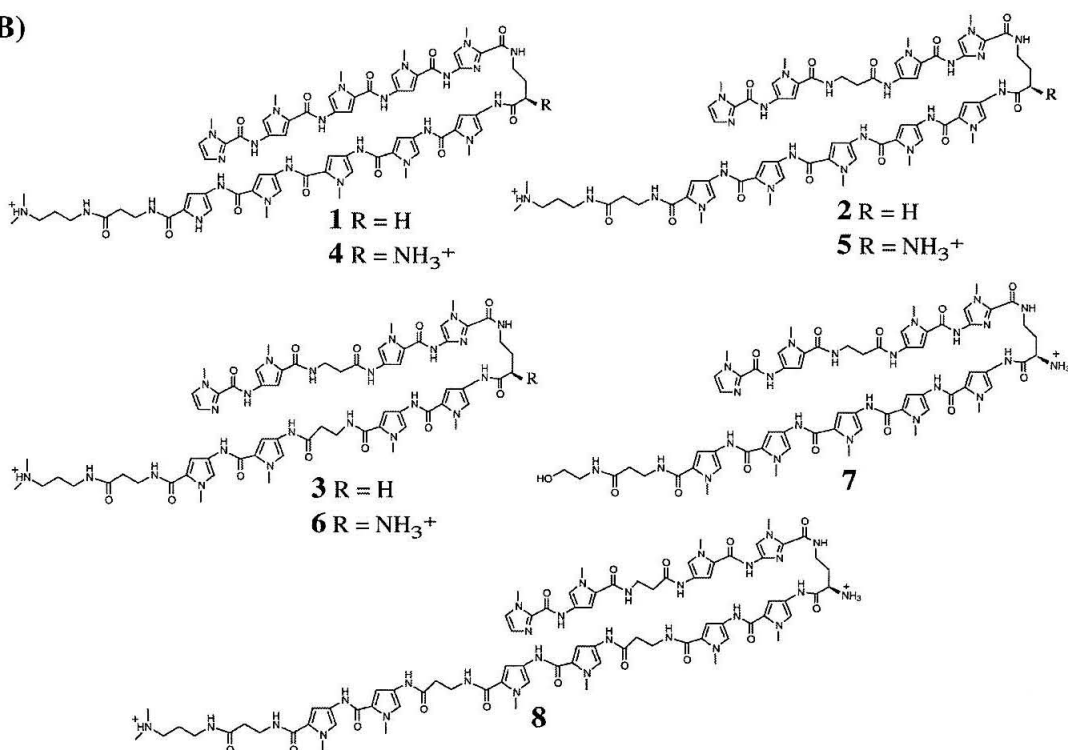
In this work, a comprehensive series of polyamides was designed to target the sequence adjacent to the HER2 TATA box on the downstream (3') side (Figures 3.3 and Figure 3.5). Polyamides targeted to this sequence, 5'-AAGAATGA-3', would also be expected to bind a similar sequence, 5'-ATGAAAGA-3', which occurs at the terminus of



the M-MuLV LTR (Chapter 2A). Quantitative DNase I footprinting titrations reveal that the same series of polyamides behaves quite differently in these two ostensibly similar contexts. For the HER2 site, high-affinity compounds were obtained which were strong inhibitors of TBP-binding and *in vitro* transcription of the HER2 gene. These studies were performed in collaboration with Joel Gottesfeld and co-workers at the Scripps Research Institute, who also investigated the *in vivo* properties of the compounds extensively in HER2-overexpressing breast cancer cells.

## Results and Discussion

**Polyamide Design.** The two G•C pairs flanked by A,T base pairs and separated by three A,T base pairs in the target site 5'-AAGAATGA-3' suggested a ten-residue hairpin polyamide<sup>37</sup> with Im/Py pairs at the first and last positions (Figure 3.5). "Ten-residue" refers to the residues in the "core" of the polyamide, not counting the C-terminal "tail" which is composed of  $\beta$ -alanine ( $\beta$ ) and *N,N*-dimethylaminopropylamine (Dp), or the linker between N- and C- terminal strands,  $\gamma$  or (*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ , which is referred to as the "turn." All of these alkyl units are specific for both A,T base pairs.<sup>38</sup> In a ten-residue polyamide, the central pair of rings can be replaced by a  $\beta/\beta$  pair (specific for both A,T base pairs), which provides increased flexibility and in many cases results in increased affinity.<sup>39,40</sup> The  $\beta$ /Py pair has also been to provide high-affinity recognition of both A,T base pairs in certain cases.<sup>41</sup> Polyamide **1** is considered the "parent" ten-ring polyamide. Polyamide **2** represents a one residue change from polyamide **1**, with the central Py/Py replaced by a  $\beta$ /Py pair. Polyamide **3** with a central  $\beta/\beta$  pair represents a two residue change from **1**.<sup>42</sup> Substitution of the  $\gamma$ -turn by the chiral, amine-functionalized derivative (*R*)<sup>H<sub>2</sub>N</sup> $\gamma$  often results in increased affinity. In a singly charged six-ring polyamide, it was shown that the (*R*)<sup>H<sub>2</sub>N</sup> $\gamma$  substitution resulted in a 10-fold gain in affinity.<sup>43</sup> In larger and/or doubly charged polyamides affinity gains of up to 10-fold have also been observed, although smaller changes are also common.<sup>44</sup> Polyamides **4-6** are the doubly charged

**Figure 3.5 A)****B)**

**Figure 3.5.** A) Ball-and-stick representations of the polyamides in this study and the target site on the HER2 promoter. The TATA box is shown as a box. Mismatches are highlighted. OH represents ethanolamine. B) Chemical structures of the designed match polyamides 1-8.

$(R)^{H_2N}\gamma$  analogs of **1-3**, respectively. Polyamide **7** is a singly charged  $(R)^{H_2N}\gamma$  analog of polyamide **5**, where Dp has been replaced by ethanolamine. Polyamide **8** is an analog of polyamide **6** which has a C-terminal extension<sup>45</sup> of  $-\beta$ -PyPy. In the target site, this C-terminal extension is placed directly in the TATA box, whereas polyamides **1-7** place only the tail inside the TATA box. Polyamides **9** and **10** are control compounds, designed as single base mismatches for the target sequence. Both resent single residue changes from polyamide **5**, with Py to Im substitutions in the C-terminal strand.

**Quantitative DNase I Footprinting.** Quantitative DNase I footprinting titrations<sup>46</sup> were performed on the series of hairpin polyamides (**1-11**) targeted to the site 5'-AAGAATGA-3' adjacent to the TATA box on the HER2 promoter, using a 188 base pair DNA fragment derived from the HER2 promoter<sup>28</sup> (Table 3.1). Representative DNase I foot-printing gels for polyamides **1** and **6** are shown in Figure 3.6, and the results of Table 3.1 are presented graphically in Figure 3.7. As a series, compounds **1-8** showed moderate to good specificity (5- to 25-fold) for the target site over mismatch sequences. Equilibrium association constants ( $K_a$ ) ranged from  $4.0 \times 10^8 \text{ M}^{-1}$  for the parent hairpin **1** to  $2.1 \times 10^{10} \text{ M}^{-1}$  for the extended hairpin **8**. More than half of this 50-fold increase in affinity for the match site occurs between the parent polyamide **1** and polyamide **6** ( $K_a = 1.2 \times 10^{10} \text{ M}^{-1}$ ), the small increase in affinity with the C-terminal extension (**8** compared to **6**) represents the rest. This small increase in affinity came with reduced specificity, as polyamide **6** showed good selectivity (>20-fold over mismatches, see Figure 3.6B), and extended hairpin **8** was the least selective (~5-fold selectivity). The ~30-fold gain in affinity for polyamide **6** over polyamide **1** appears to be the result of both substitution of the central Py/Py pair to  $\beta/\beta$  and substitution of the  $\gamma$ -turn to the  $(R)^{H_2N}\gamma$ -turn. Both changes to polyamide **1** resulted in an ~10-fold increase in affinity (polyamides **3** and **4**), and then the complimentary change resulted in a further ~3-fold affinity gain (**6** as compared to **3,4**).

**Table 3.1.** Equilibrium association constants for the site 5'-atAGAATGAag-3' adjacent to the TATA box on the HER2 promoter.<sup>a</sup>

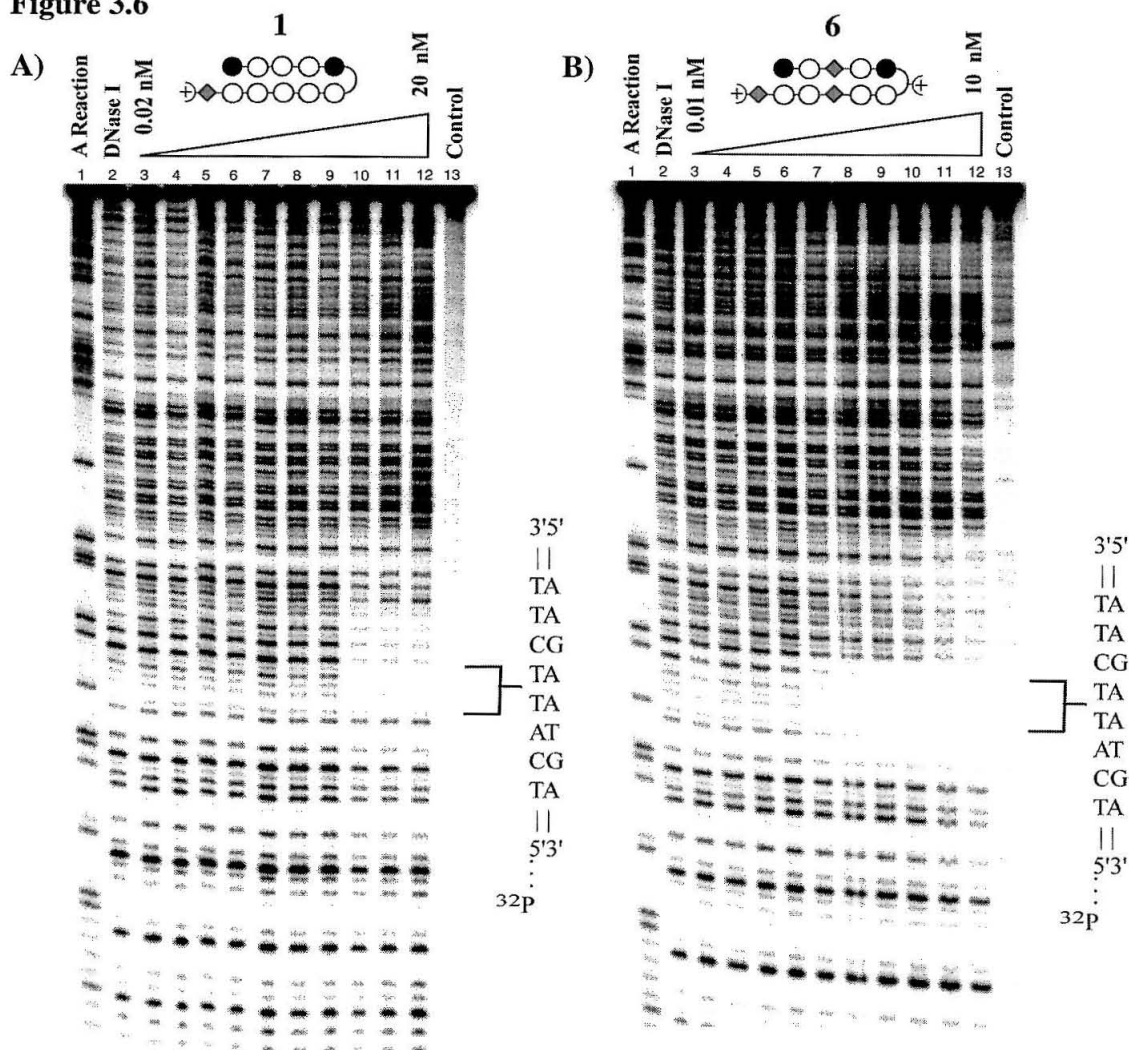
Polyamide	$K_a$ (M <sup>-1</sup> )
1	$4.0 \times 10^8$
2	$5.5 \times 10^8$
3	$4.8 \times 10^9$
4	$4.3 \times 10^9$
5	$3.8 \times 10^9$
6	$1.2 \times 10^{10}$
7	$2.5 \times 10^9$
8	$2.1 \times 10^{10}$
9	$7.2 \times 10^8$
10	$6.9 \times 10^8$

**Table 3.1.** Values reported are the mean values obtained from at least three independent DNase I footprint titration experiments. The assays were carried out at 22° C, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. (a) Polyamide 8 recognizes the larger binding site, 5'-taTATAAGAATGAag-3'.

The central  $\beta$ /Py pair compounds were of similar affinity to the central Py/Py pair polyamides and lower affinity than the central  $\beta$ / $\beta$  pair polyamides having the same turn and tail ( $1 \approx 2 < 3$ ;  $4 \approx 5 < 6$ ). Based on the central  $\beta$ /Py pair compounds, the affinity increases observed for the (*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ -turn are due not only to the resulting polyamides being doubly charged, but also to a specific gain for the placement of an amine group in a chiral position on the  $\gamma$ -turn: singly charged (*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ -containing compound 7 shows a 4.5-fold increase in affinity from 2, more than half of the 7-fold affinity increase for 2 to 5. The two control compounds 9 and 10 also have a central  $\beta$ /Py pair, and, while their remarkably similar affinities are 2-fold lower than 5, they are actually greater than 2 by a small margin.

Relatively strong binding of mismatch compounds is generally not desired, but it is potentially advantageous for polyamide 9 on the HER2 promoter. This is because there is a cognate site (5'-CTGCTTGA-3', not a perfect match site due to a Dp mismatch, underlined) for 9 adjacent to the Ets site (bound by PA-1 in Figure 3.3). Polyamide 9

Figure 3.6



**Figure 3.6.** Quantitative DNase I footprint titration experiments on a 5'-<sup>32</sup>P-labeled 188-bp DNA fragment of the HER2 promoter.<sup>28</sup> All reactions contain 15 kcpm DNA fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. **A)** Lane 1, A reaction; lane 2, DNase I standard; lanes 3-12, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, and 20 nM polyamide 1, respectively; lane 13, intact DNA. **B)** Lane 1, A reaction; lane 2, DNase I standard; lanes 3-12, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide 6, respectively; lane 13, intact DNA.

binds this site with 2-fold higher affinity ( $K_a = 1.2 \times 10^9 \text{ M}^{-1}$ ) than the 5'-AAGAATGA-3' site (mismatch underlined). The relatively similar affinity for both sites suggests that polyamide 9 may be able to interfere with the binding of both transcription factors, which could be beneficial for transcription inhibition.

**Table 3.2.** Equilibrium association constants for the M-MuLV DNA fragment.

Polyamide/Binding Site	$K_a$ (M <sup>-1</sup> )	Polyamide/Binding Site	$K_a$ (M <sup>-1</sup> )
<b>1</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$2.2 \times 10^9$	<b>6</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$1.6 \times 10^9$
<b>2</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$4.8 \times 10^8$	<b>7</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$1.9 \times 10^9$
<b>3</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$1.8 \times 10^9$	<b>8</b> 5'-G G G A A T G A A A G A C C-3' 3'-C C C T T A C T T T C T G G-5' 	$2.7 \times 10^9$
<b>4</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$2.5 \times 10^9$	<b>9</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$2.7 \times 10^7$
<b>5</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$3.6 \times 10^9$	<b>10</b> 5'-G A A T G A A A G A C C-3' 3'-C T T A C T T T C T G G-5' 	$2.3 \times 10^7$

**Table 3.2.** Values reported are the mean values obtained from at least three independent DNase I footprint titration experiments. The assays were carried out at 22° C, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. Mismatches are highlighted in grey.

**Comparison with the site 5'-ATGAAAGA-3' from the M-MuLV LTR.** The effects of  $\beta$ -alanine substitutions,<sup>39-41</sup> the  $\gamma$ -turn or (*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ -turn,<sup>43</sup> and C-terminal extensions<sup>45</sup> have all been studied previously, but have typically been considered separately. Rarely have these substitutions been considered together for a series of polyamides targeting two different binding sites on two different DNA fragments. In addition to the HER2 site, 5'-AAGAATGA-3', polyamides **1-10** were also investigated on a second binding site, 5'-ATGAAAGA-3', which occurs near the terminus of the M-MuLV LTR (Chapter 2A). Quantitative DNase I footprinting titrations with a 183 base pair DNA fragment containing an internal copy of the M-MuLV LTR were performed as described in Chapter 2A. (For a representative DNase I footprinting gel, see Figure 2.2A, compound “2” in Chapter 2A is polyamide **5** in this chapter). The equilibrium

Figure 3.7

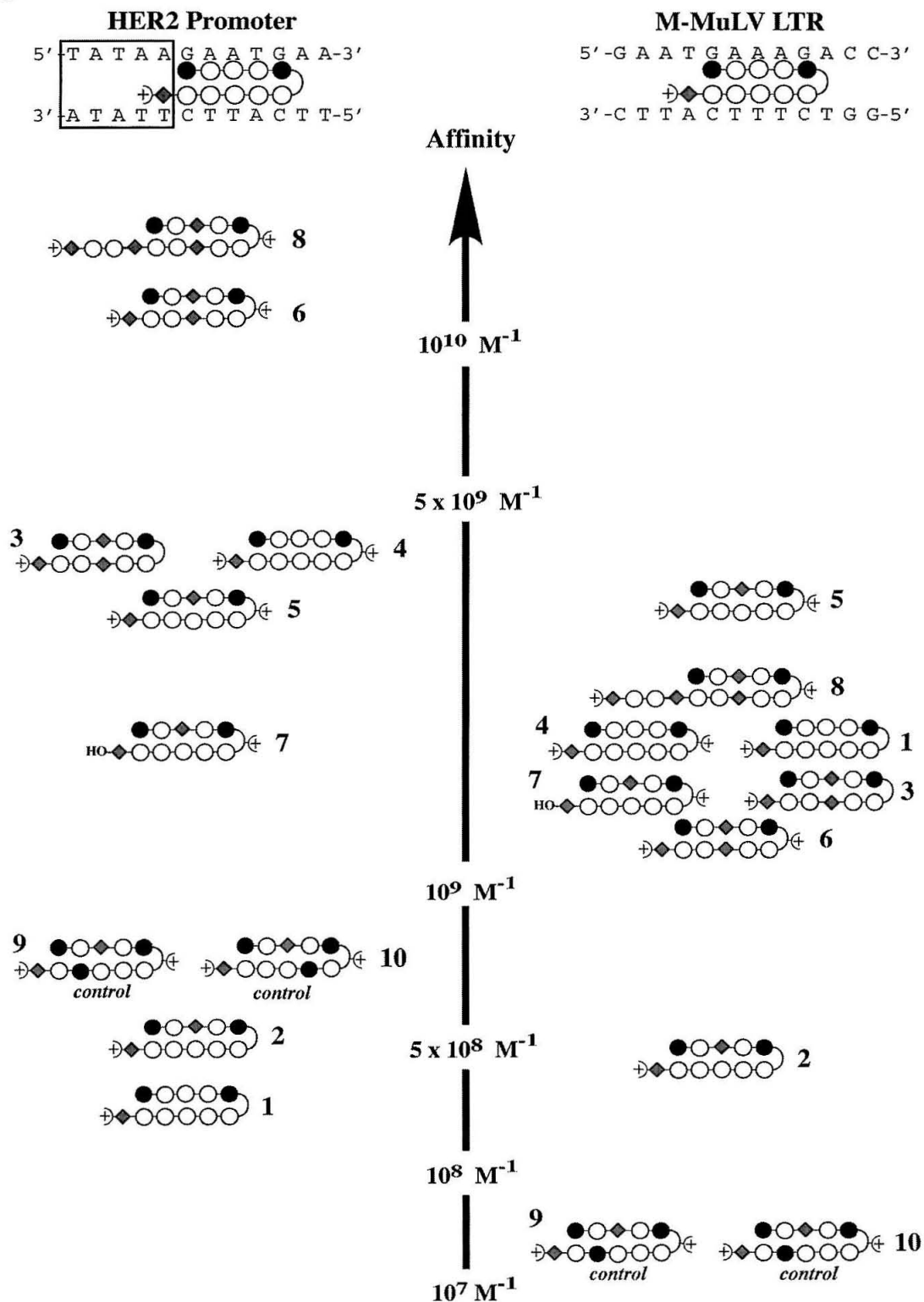


Figure 3.7. Comparison of polyamides 1-10 for two similar binding sites.

association constants of polyamides **1-10** on the M-MuLV fragment are listed in Table 3.2 and a comparison with the results from the HER2 fragment (Table 3.1) is shown in Figure 3.7. Compared to the HER2 site, the range of affinities for the M-MuLV site is much tighter, with seven of the eight designed match polyamides clustering between  $K_a = 1.6 \times 10^9 \text{ M}^{-1}$  (**6**) and  $K_a = 3.6 \times 10^9 \text{ M}^{-1}$  (**5**). It should be noted that extended hairpin **8** places the  $\beta$ -Dp tail against G•C base pairs, so this site is not a fair comparison to the HER2 site for **8**. All of the compounds were more specific for the match site on the HER2 fragment than the M-MuLV fragment, where specificity for the target site was generally less than 2-fold. Polyamides with an Im/Py pair adjacent to the turn (either  $\gamma$  or  $(R)^{H_2N}\gamma$ ) including **1-8** on the M-MuLV fragment, have shown a high propensity to bind A,T base pairs as well as G•C at that position (in contrast to a Py/Im pair adjacent to the turn, which is highly specific for C•G). The specificity of this series for the target site on the HER2 fragment may be due to the fact that this common mismatch is under-represented on this heavily G,C-rich region of the HER2 promoter (Figure 3.2).

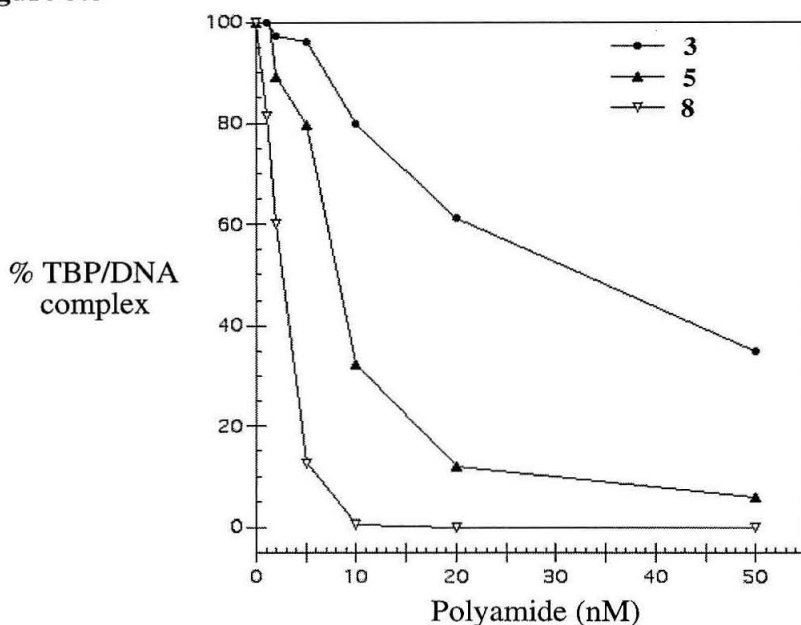
It is interesting that the compounds that are the most similar in absolute affinity between the two sites are the central  $\beta$ /Py pair match polyamides (**2,5,7**), while the central  $\beta$ /Py pair mismatch polyamides (**9,10**) bind the M-MuLV site with 30-fold lower affinity than the HER2 site. Two other compounds which vary significantly between the two sites are the parent ten-ring polyamide **1** and the central  $\beta$ / $\beta$  pair,  $(R)^{H_2N}\gamma$ -turn containing **6**. The 30-fold difference in favor of **6** for the HER2 site has vanished. Whereas for the HER2 site parent polyamide **1** had lower affinity than the control compounds, for the M-MuLV site it has 100-fold higher affinity than the controls. The affinity of polyamide **1** is 5.5-fold higher for the M-MuLV site than for the HER2 site, but this affinity was not improved by single  $\beta$ -alanine substitution (**2**, a 4.5-fold drop from **1**), the central  $\beta$ / $\beta$  pair (**3**  $\approx$  **1**), or the  $(R)^{H_2N}\gamma$ -turn (**4**  $\approx$  **1**). While the  $(R)^{H_2N}\gamma$ -turn did not effect the central  $\beta$ / $\beta$  pair compounds (**6**  $\approx$  **3**) either, it did have a significant effect on



the central  $\beta$ /Py pair compounds, and as for HER2, the affinity increase (7.5-fold, **2** to **5**) was not solely due to the presence of two charges (**2** to **7**, 4-fold increase).

The differences observed between these two ostensibly similar sites are likely the result of structural differences between the sites themselves, differences in the flanking sequences, and differences between the DNA fragments as a whole. Both sites are largely homopurine/homopyrimidine tracts and contain 5'-GA-3' steps, which are known to be difficult for polyamide recognition.<sup>44,47</sup> Despite or perhaps because of this, the parent polyamide **1** appears to be especially well matched for the M-MuLV sequence, as changes in the structure failed to improve binding significantly. The parent polyamide **1** was disfavored by the HER2 site, even compared to the control polyamides (**9,10**). However, changes in structure produced up to 50-fold higher affinity binders, with the highest affinity compounds (**6,8**) being ~5-fold tighter binders than the highest affinity compound for the M-MuLV site (**5**). While central  $\beta$ / $\beta$  pair and (*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ -turn substitutions are not general solutions for all homopurine/homopyrimidine tracts with 5'-GA-3' steps, the affinity increases afforded by these structural alterations for the 5'-AAGAATGA-3' site generated good candidates for the inhibition of HER2 transcription.

**Inhibition of TBP-binding and HER2 *in vitro* transcription.** Selected polyamides were tested for the inhibition of TBP-binding to the TATA box of the HER2 promoter in electrophoretic mobility shift assays (Figure 3.8). Polyamide **3** displayed an IC<sub>50</sub> (the concentration required for 50% inhibition of TBP-binding) between 20 and 50 nM. A control single base mismatch compound (not shown in Figure 3.8) showed 25% inhibition at 50 nM. Polyamides **5** and **8** were extremely potent inhibitors. The IC<sub>50</sub> between 5 and 10 nM for polyamide **5** is surprising compared to the value for **3**, given the similar, and in fact slightly higher, affinity of **3** ( $K_a = 4.8 \times 10^9 \text{ M}^{-1}$  compared to **5**,  $K_a = 3.8 \times 10^9 \text{ M}^{-1}$ ). The extended hairpin **8** was the most effective, yielding 40% inhibition at 2 nM and 85% inhibition at 5 nM. TBP-binding was completely inhibited at 10 nM.

**Figure 3.8****Figure 3.8.** Inhibition of TBP binding to the HER2 promoter in electrophoretic mobility shift assays.

The potent inhibitors of TBP-binding were tested in *in vitro* transcription assays with the HER2 promoter. Polyamide **5** showed 70% inhibition at 1  $\mu\text{M}$  ( $\text{IC}_{50} \approx 600 \text{ nM}$ ) and extended hairpin **8** showed complete inhibition at 1  $\mu\text{M}$  with an  $\text{IC}_{50}$  value just above 100 nM (45% inhibition at 100 nM). These values are quite promising for *in vitro* transcription experiments with the HER2 promoter,<sup>28</sup> and all further efforts with polyamides **1-10** were concentrated on *in vivo* assays. However, *in vitro* assays with **6** and **7** would be useful. The similar affinity and structure of polyamide **6** to the extended hairpin **8** could be used to gage whether the extremely potent inhibition demonstrated by **8** is due to high affinity alone ( $K_a = 2.1 \times 10^{10} \text{ M}^{-1}$ ) or if the C-terminal extension into the TATA box is particularly beneficial. The higher efficacy of **5** as compared to **3** may be due to the former being doubly charged, or perhaps the position of the amino group on the  $(R)^{\text{H}_2\text{N}}\gamma$ -turn of **5** in the minor groove plays a role. These possibilities could be partially distinguished by polyamide **7**. These additional studies would be informative

for the design of future polyamide inhibitors of TBP-binding. Additionally, based on affinity, polyamide **6** should be a strong inhibitor of HER2 transcription.

**Effect of polyamides on HER2 transcription in HER2-overexpressing breast cancer cells.** The ultimate goal of this project has been the inhibition of transcription of an endogenous human gene in human cells. Towards that end polyamides from this series, particularly **3**, **5**, and **8**, have been studied extensively in a variety of different *in vivo* assays with SKBR-3 and MCF-7 HER2-overexpressing human breast cancer cells. These include RT-PCR analysis of the HER2 mRNA, RNase protection assays with a probe specific for the HER2 mRNA, fluorescence-activated cell sorting (FACS) analysis of the HER2 protein with an anti-HER2 antibody, and cell motility, which has been shown to be a function of the amount of HER2 protein present in the membrane of HER2-overexpressing cancer cells such as SKBR-3. In all cases except FACS (which, in retrospect may be the most reliable assay), initial positive results were obtained which did not prove to be reproducible. In the case of the mRNA-based assays, the initial results were extremely positive with  $IC_{50}$  values less than 50 nM obtained for polyamides **5** and **8**. Control experiments using probes for eight other genes in the RNase protection assay showed that these mRNA levels were not significantly affected by **5**, and SKBR-3 cells were shown to be viable at up to 3  $\mu$ M **5**. Since down-regulation of HER2 in SKBR-3 cells often results in cell-cycle arrest and apoptosis,<sup>22,23,27</sup> the cells probably should not have maintained viability with 3  $\mu$ M of a 50 nM inhibitor. Indeed, further RNase protection assays showed no inhibition by **5** or other polyamides. Selected polyamides from the **1-10** series were tested with FACS analysis of the HER2 protein levels after treatment with polyamide, both alone and in combination with ESX-binding inhibitors of HER2 *in vitro* transcription such as **PA-1**,<sup>28</sup> and no inhibition was observed.

Concurrent to these efforts in the Gottesfeld group (FACS analysis and cell motility in collaboration with Dr. Susann Shenk and Professor Vito Quaranta), the

Beerman group at Roswell Park Cancer Institute was investigating the effects of the ESX-binding inhibitors<sup>28</sup> and AP-2 binding inhibitors (see Figure 3.3) on SKBR-3 cells. No inhibition was observed *in vivo*. It was hypothesized that polyamides may have poor uptake properties in SKBR-3 cells. Surprisingly, polyamide-fluorescent dye conjugates were permeable to live SKBR-3 cells but did not localize to the nucleus (see Chapter 4A). Based on the dye-conjugates, the failure of inhibition *in vivo*, and other supportive indirect evidence, hairpin polyamides such as **1-10** probably do not localize to the nucleus of SKBR-3 or MCF-7 cells.

## Conclusions and Outlook

A series of ten polyamides including two mismatch controls was designed to bind to the sequence 5'-AAGAATGA-3' adjacent to the TATA box of the HER2 promoter. A variety of substitutions to the parent ten-ring polyamide structure produced high-affinity binders for this site. The effect of these substitutions was further tested by targeting a closely related sequence 5'-ATGAAAGA-3' derived from the M-MuLV LTR. The parent ten-ring polyamide was especially well matched for the latter sequence, as the structural alterations which produced a 30-fold increase in affinity for the HER2 site showed little effect. Polyamides targeted to the HER2 site were potent inhibitors of TBP-binding to the TATA box and *in vitro* transcription of the HER2 gene. However, extensive experiments in cell culture have shown that these potent *in vitro* inhibitors are not effective *in vivo*. It is likely that the hairpin polyamides used in this study do not localize to the nucleus of the breast cancer cells examined thus far. Efforts are underway to screen additional breast cancer cell-lines, find reagents to allow normal hairpin polyamides to reach the nucleus, and produce modified polyamides with enhanced nuclear localization properties (Chapter 4). When these efforts are successful, polyamides from this series, or modified polyamides based on the lessons learned from this series, may prove to be potent inhibitors of HER2 transcription in human cells.

## Appendix: Blocking Transcription Through a Nucleosome\*

\*Gottesfeld, J.M.; Belitsky, J.M.; Melander, C.; Dervan, P.B.; Luger, K. "Blocking Transcription Through a Nucleosome with Small DNA Ligands." *J. Mol. Biol.* submitted 2002.

One of the compounds from the series designed to bind adjacent to the TATA box in the HER2 promoter was used by Joel Gottesfeld in an exciting demonstration of transcription inhibition not directly related to HER2.\* Previous studies have established that most sites on nucleosomal DNA are accessible to polyamides, and that nucleosomes remain fully folded upon ligand binding. Polyamide **3** was one of four polyamides targeted to a DNA fragment containing the sea urchin 5S gene nucleosome positioning sequence linked to a T7 RNA polymerase promoter. Transcription of this construct in the presence and absence of polyamides and pre-formed histone octamers. Polyamide **3** and another polyamide bound within the sea urchin 5S gene nucleosome positioning sequence and inhibited both heat-induced nucleosome sliding and transcription by T7 RNA polymerase from the nucleosomal template but not from histone-free DNA. These polyamides also prevented repositioning of the histone octamer by RNA polymerase, thereby inhibiting passage of the elongating polymerase through nucleosomal DNA. The results establish the requirement for octamer mobility for transcription of nucleosomal templates by T7 RNA polymerase. The effect of the polyamides may be due to the prevention of rotation of the DNA on the surface of the histone octamer. This may ultimately provide a means of transcription inhibition that does not depend on competing with transcription factors and may be amenable to inhibition from sites within the coding region of genes, rather than being limited to promoter sites.

## Experimental

**Polyamides.** All polyamides were synthesized by solid-phase methods as previously described,<sup>43,48</sup> and characterized by analytical HPLC and MALDI-TOF mass

spectrometry. ImPyPyPyIm-(R)- $\gamma$ -PyPyPyPyPy- $\beta$ -Dp (1) [M+H] 1466.99, 1466.57 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)- $\gamma$ -PyPyPyPyPy- $\beta$ -Dp (2) [M+H] 1415.66, 1415.65 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)- $\gamma$ -PyPy- $\beta$ -PyPy- $\beta$ -Dp (3) [M+H] 1364.88 1364.64 calculated for [M+H]. ImPyPyPyIm-(R)- $^{H^{2N}}\gamma$ -PyPyPyPyPy- $\beta$ -Dp (4) [M+H] 1481.98, 1481.68 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)- $^{H^{2N}}\gamma$ -PyPyPyPyPy- $\beta$ -Dp (5) [M+H] 1430.94, 1430.67 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)- $^{H^{2N}}\gamma$ -PyPy- $\beta$ -PyPy- $\beta$ -Dp (6) [M+H] 1379.91, 1379.65 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)- $^{H^{2N}}\gamma$ -PyPyPyPyPy- $\beta$ -EtOH (7) [M+H] 1389.63, 1389.60 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)- $^{H^{2N}}\gamma$ -PyPy- $\beta$ -PyPy- $\beta$ -PyPy- $\beta$ -Dp (8) [M+H] 1695.08, 1694.79 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)- $^{H^{2N}}\gamma$ -PyPyPyImPy- $\beta$ -Dp (9) [M+H] 1431.84, 1431.68 calculated for [M+H]. ImPy- $\beta$ -PyIm-(R)- $^{H^{2N}}\gamma$ -PyImPyPyPy- $\beta$ -Dp (10) [M+H] 1431.48, 1431.68 calculated for [M+H].

**DNase I footprinting.** DNase I footprinting reactions were performed as previously described<sup>47</sup> for  $^{32}\text{P}$ -labeled DNA fragment derived from the HER2 promotor<sup>28</sup> and an internal copy of the M-MuLV LTR (Chapter 2A).

***In vitro* assays.** Electrophoretic mobility shift assays measuring polyamide inhibition of TBP binding to a DNA fragment derived from the HER2 promoter and *in vitro* transcription assays with the HER2 promoter were performed according to standard methods as previously described.<sup>28,33</sup>

## Acknowledgements

We are grateful to the NIH for research support and to the Ralph M. Parsons foundation for a predoctoral fellowship to J.M.B. We thank Roland Burli, Christian Melander, Joel Gottesfeld, and Terry Beerman for many thoughtful discussions. We are grateful to Eldon Baird and John Trauger for early work on this project and acknowledge

co-workers in the Gottesfeld, Beerman, and Quaranta labs, including (but not limited to) Stephanie Leslie, Christine White, Susann Shenk, John Long, and Laura Neely for their hard work.

## References and Notes

- (1) Slamon, D. J.; Clark, G. M.; Wong, S. G.; Levin, W. J.; Ullrich, A.; McGuire, W. L. *Science* **1987**, *235*, 177-182.
- (2) Slamon, D. J.; Jones, L. A.; Holt, J. A.; Wong, S. G.; Keith, D. E.; Levin, W. J.; Stuart, S. G.; Udove, J.; Ullrich, A. *Science* **1989**, *244*, 707-712.
- (3) Singleton, T. P.; Strickler, J. G. *Pathol. Annu.* **1992**, *27*, 165-190.
- (4) Tzahar, E.; Yarden, Y. *BBA Rev. Cancer* **1998**, *1377*, M25-M37.
- (5) King, C. R.; Kraus, M. H.; Aaronson, S. A. *Science* **1985**, *229*, 974-976.
- (6) Benz, C. C.; O' Hagan, R. C.; Richter, B.; Scott, G. K.; Chang, C.-H.; Xiong, X.; Chew, K.; Ljung, B.-M.; Edgerton, S.; Thor, A.; Hassell, J. A. *Oncogene* **1997**, *15*, 1513-1525.
- (7) Stotelers, C.; van de Poll, M. L. M.; Lenferink, A. E. G.; Gadellaa, M. M.; van Zoelen, C.; van Zoelen, E. J. J. *Biochemistry* **2002**, *41*, 4292-4301.
- (8) Wilkenson, J. C.; Stein, R. A.; Guyer, C. A.; Beechem, J. M.; Staros, J. V. *Biochemistry* **2001**, *40*, 10230-10242.
- (9) Wilkenson, J. C.; Staros, J. V. *Biochemistry* **2002**, *41*, 8-14.
- (10) Sharpe, S.; Barber, K. R.; Grant, C. W. M. *Biochemistry* **2002**, *41*, 2341-2352.
- (11) Watabe, T.; Yoshida, K.; Shindoh, M.; Kaya, M.; Fujikawa, K.; Sato, H.; Seiki, M.; Ishii, S.; Fujinaga, K. *Int. J. Cancer* **1998**, *77*, 128-137.
- (12) O' Hagan, R. C.; Hassell, J. A. *Oncogene* **1998**, *16*, 301-310.
- (13) Galang, C. K.; Garcia-Ramirez, J. J.; Solski, P. A.; Westwick, J. K.; Der, C. J.; Neznanov, N. N.; Oshima, R. G.; Hauser, C. A. *J. Biol. Chem.* **1996**, *273*, 7992-7998.

- (14) Chang, C. H.; Scott, G. K.; Kuo, W. L.; Xiong, X.; Suzdaltseva, Y.; Park, J. W.; Sayre, P.; Erny, K.; Collins, C.; Gray, J. W.; Benz, C. C. *Oncogene* **1997**, *14*, 1617-1622.
- (15) Oettgen, P.; Barcinski, M.; Boltax, J.; Stolt, P.; Akbarali, Y.; Liberman, T. A. *Genomics* **1999**, *55*, 358-362.
- (16) McKeage, K.; Perry, C. M. *Drugs* **2002**, *62*, 209-243.
- (17) Vogel, C. L.; Cobleigh, M. A.; Tripathy, D.; Gutheil, J. C.; Harris, L. N.; Fehrenbacher, L.; Slamon, D. J.; Murphy, M.; Novotny, W. F.; Burchmore, M.; Shak, S.; Stewart, S. J.; Press, M. J. *Clin. Oncol.* **2002**, *20*, 719-726.
- (18) Neve, R. M.; Nielsen, U. B.; Kirpotkin, D. B.; Poul, M.-A.; Marks, J. M.; Benz, C. C. *Biochem. Biophys. Res. Commun.* **2001**, *280*, 274-279.
- (19) Berezov, A.; Zhang, H.-T.; Greene, M. I.; Murali, R. J. *Med. Chem.* **2001**, *44*, 2565-2574.
- (20) Correa, I.; Plunkett, T. *Breast Cancer Res.* **2001**, *3*, 399-403.
- (21) Hurst, H. C. *Breast Cancer Res.* **2001**, *3*, 395-398.
- (22) Roh, H.; Pippin, J. A.; Drebin, J. A. *Cancer Res.* **2000**, *60*, 560-565.
- (23) Roh, H.; Pippin, J. A.; Green, D. W.; Boswell, C. B.; Hirose, C. T.; Mokadam, N.; Drebin, J. A. *Oncogene* **2000**, *19*, 6138-6143.
- (24) Hudson, L. G.; Ertl, A. P.; Gill, G. N. *J. Biol. Chem.* **1990**, *265*, 4389-4393.
- (25) Porumb, H.; Gousset, H.; Letellier, R.; Salle, V.; Briane, D.; Vassy, J.; Amor-Gueret, M.; Israel, L.; Taillandier, E. *Cancer Res.* **1996**, *56*, 515-522.
- (26) Beerli, R. R.; Segal, D. J.; Dreier, B.; Barbas, C. F., III *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14628-14633.
- (27) Beerli, R. R.; Dreier, B.; Barbas, C. F., III *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1495-1500.
- (28) Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A. *J. Biol. Chem.* **2000**, *275*, 24246-24254.



- (29) Gottesfeld, J. M.; Turner, J. M.; Dervan, P. B. *Gene Expression* **2000**, *9*, 77-91.
- (30) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.
- (31) Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688-693.
- (32) Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. *J. Mol. Biol.* **2001**, *309*, 615-629.
- (33) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895.
- (34) Burli, R.W.; Dervan, P.B.; Beerman, T.A. Unpublished results.
- (35) Burley, S. K. *Curr. Opin. Struct. Biol.* **1996**, *6*, 69-75.
- (36) Lewin, B. *Genes V*; Oxford University Press: New York, 1994.
- (37) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636-7644.
- (38) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.
- (39) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166.
- (40) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.
- (41) Wang, C. C. C.; Ellervik, U. *Bioorg. Med. Chem.* **2001**, *9*, 653-657.
- (42) Polyamide **3** was originally synthesized by Eldon Baird, and subsequently by the author.
- (43) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.
- (44) Dervan, P. B. Unpublished results.
- (45) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1996**, *3*, 369-377.
- (46) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.

- (47) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532-12537.
- (48) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

## **CHAPTER 4**

### **Cellular Uptake and Localization of DNA Binding Polyamide-Fluorescent Dye Conjugates**

*The text of this chapter was taken in part from a manuscript submitted for publication coauthored with Paramjit Arora and Professor Peter Dervan (Caltech) and Stephanie Leslie and Professor Terry Beerman (Roswell Park Cancer Institute).*

(Belitsky, J.M.; Leslie, S.J.; Arora, P.S.; Beerman, T.A.; Dervan, P.B. "Cellular Uptake of *N*-Methylpyrrole/*N*-Methylimidazole Polyamide-Dye Conjugates." *Bioorg. Med. Chem*, **in press**.)

## Abstract

The cellular uptake and localization properties of DNA binding *N*-methylpyrrole/*N*-methylimidazole polyamide-dye conjugates in a variety of living cells have been examined by confocal laser scanning microscopy. Chapter 4A provides an introduction to these studies and our initial results. With the exception of certain T-cell lines, polyamide-dye conjugates localize mainly in the cytoplasm and not in the nucleus. Reagents such as methanol typically used to fix cells for microscopy significantly alter the cellular localization of these DNA-binding ligands.

Chapter 4B describes confocal microscopy studies of modified polyamide-dye conjugates in a wide variety of cell lines. Different motifs such as the cycle, tandem, and one-to-one binding polyamide-dye conjugates have been examined, and point to molecular weight as an important factor for nuclear localization in some cell lines. Modifications include the alkylating agent chlorambucil, carrier peptides such as TAT, and small molecules aimed at receptor-mediated uptake. Of these, cationic peptide conjugates have proven the most useful. However, these compounds show poor DNA binding properties. Efforts to balance nuclear localization with favorable DNA binding specificity are discussed. Multiple solutions for several cell lines have emerged, and on-going experiments should define further solutions.

**Chapter 4A:**  
**Cellular Uptake of *N*-Methylpyrrole/*N*-Methylimidazole**  
**Polyamide-Dye Conjugates**

## **Introduction**

Cell-permeable small molecules with the ability to target predetermined DNA sequences and interfere with gene expression would be useful tools in molecular biology and, potentially, human medicine. Polyamides containing aromatic amino acids *N*-methylpyrrole (Py), and *N*-methylimidazole (Im) are synthetic ligands that have an affinity and specificity for DNA comparable to naturally occurring DNA-binding proteins.<sup>1,2</sup> Based on pairing rules for recognition in the minor groove, polyamides can be designed to target a large number of predetermined DNA sequences. Polyamides have been shown to inhibit several classes of transcription factors, and thus regulate transcription in cell-free systems.<sup>3</sup> Modification of the DNA binding polyamide scaffold has produced agents capable of sequence-specific DNA-alkylation,<sup>4,5</sup> recruitment of Topoisomerase I,<sup>6</sup> and gene activation.<sup>7,8</sup> Significantly, polyamides have recently been shown to bind specific DNA sequences in isolated nucleosomes.<sup>9</sup>

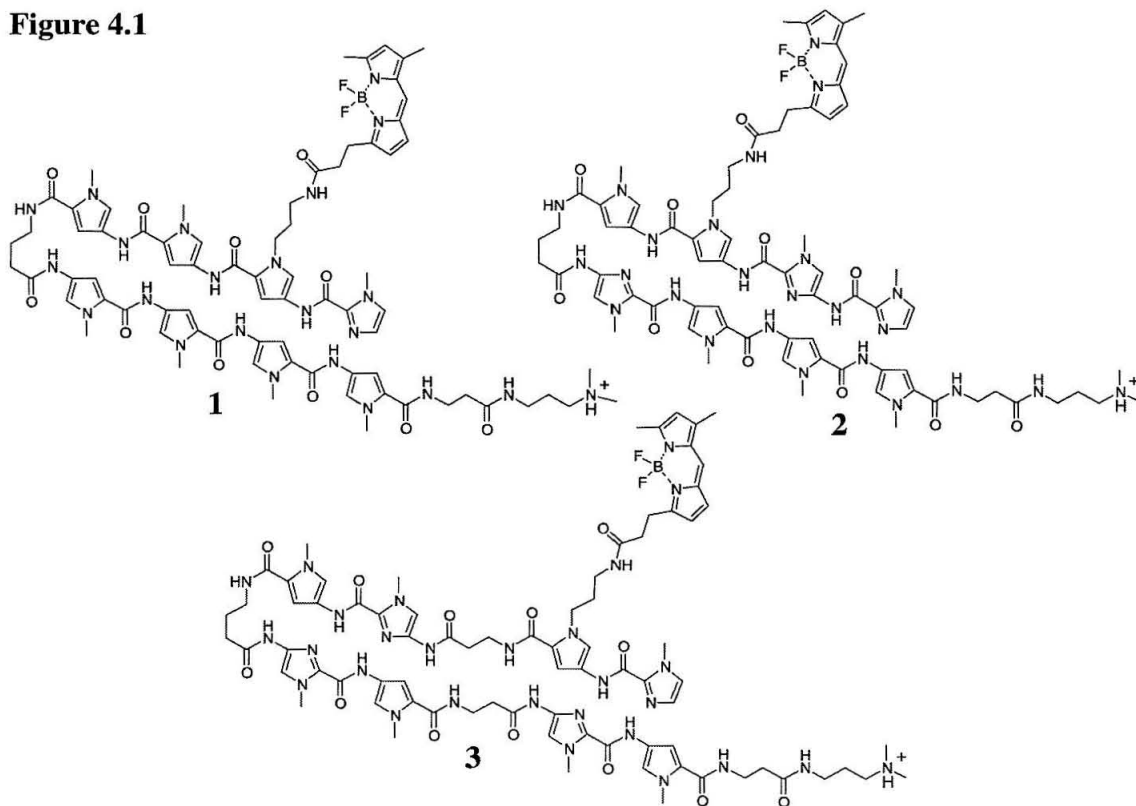
Current impetus in our laboratories is directed at evaluating the potential of these molecules as potent regulators of specific gene transcription in cell culture assays. Polyamides targeted to the HIV-1 promoter inhibited viral replication in human peripheral blood lymphocytes.<sup>10</sup> Laemmli and coworkers have shown that polyamides targeting heterochromatic satellite DNA induced specific gain- and loss-of-function phenotypes in *Drosophila* presumably as a result of chromatin opening.<sup>11,12</sup> Together these studies indicated that polyamides are cell-permeable and, assuming that DNA is the cellular target of polyamides, localize to the nucleus at least in some cases.

In order to determine the potential of polyamides to inhibit the transcription of endogenous genes in human cells, our initial studies focused on the *HER2/neu* gene.

Overexpression of the HER2/*neu* gene is implicated in human breast cancer and associated with poor prognosis.<sup>13,14</sup> Hairpin polyamides targeted to the Ets binding site of the HER2/*neu* promoter were shown to inhibit binding of the transcription factor ESX at nanomolar concentrations and to inhibit transcription of the HER2/*neu* gene in cell-free experiments.<sup>15</sup> However, despite the demonstrated *in vitro* activity, no transcription inhibition was observed in HER2-overexpressing SKBR-3 cells.<sup>16</sup> In order to test whether this result might be due to poor cellular uptake or nuclear localization, a series of polyamide conjugates incorporating the fluorophore Bodipy FL were synthesized. Confocal microscopy was used to study the cellular distribution of polyamide-Bodipy conjugates in SKBR-3 cells. In initial microscopy studies, cells were incubated with polyamide-Bodipy conjugates and visualized after treatment with cell-fixing reagents. Polyamide-dye conjugates were observed to localize in the nucleus of fixed cells. However, it was discovered that the common procedure of fixing cells prior to imaging had a drastic effect on conjugate localization, and that in living SKBR-3 cells without fixation, nuclear localization was not observed within detection limits. This suggested that a study of the cellular localization of polyamides in a variety of living cells was warranted. Here we report our initial findings on this key issue for gene regulation by polyamides.

## Results

**Polyamide-Bodipy Conjugates.** Bodipy FL was chosen as a fluorophore because of its enhanced photostability relative to fluorescein and its relatively small size.<sup>17</sup> The fluorescence of Bodipy is independent of pH in the physiological range, and the free dye is not known to localize to the nuclei of cells which should limit false positive results. Three polyamide-Bodipy FL conjugates, eight-ring hairpins **1** and **2** and a 2- $\beta$ -2 hairpin (**3**) were synthesized (Figure 4.1). Polyamides **1** and **2** are the same size but differ in sequence and the number of Im/Py rings. The polyamide precursors were

**Figure 4.1****Figure 4.1.** Polyamide-Bodipy FL conjugates.

synthesized by solid-phase methods<sup>18</sup> with one Py residue modified to yield a free amine upon cleavage from the resin.<sup>19</sup> The free amine of the precursor polyamides was allowed to react with the succinimidyl ester of Bodipy FL (Molecular Probes) affording polyamide conjugates **1-3**. The purity and identity of the dye conjugates were verified by analytical HPLC, UV-Vis spectroscopy, and MALDI-TOF mass spectrometry. Additionally, DNase I footprinting experiments<sup>20</sup> were performed to assay the ability of the conjugates to bind DNA. Despite the short tether between the bulky dye and the polyamide backbone, the Bodipy conjugates still maintain high affinity and specificity for their match sites ( $K_a = 10^7$ - $10^9$  M<sup>-1</sup>) although with a loss of 10- to 100-fold in affinity relative to their parent polyamide. The fluorescence of conjugates was unaffected by the presence of cognate DNA or acidic pH. There was a 2-fold increase in fluorescence in methanol as compared to buffer.

**Confocal Microscopy.** Confocal microscopy was used to study the distribution of polyamide-Bodipy conjugates in SKBR-3 cells. Using a procedure that included fixing the cells in methanol before imaging, rapid uptake and nuclear localization was observed.<sup>21</sup> However, in a control experiment without fixation, significantly reduced fluorescence located in a punctate pattern in the cytoplasm was observed (Figure 4.2A). With the observation that methanol influences cellular localization of polyamides, an effort to screen a variety of *live* cells with the Bodipy-polyamide conjugates was initiated. Cells were incubated with 5  $\mu$ M concentration of conjugate **1-3** or free Bodipy FL as control in a total volume of 100  $\mu$ L for 20 hours under normal growth conditions for each cell line (see Legend, Table 4.1). Most incubations were performed in culture dishes equipped with glass bottoms for direct imaging (MatTek Corporation). Adherent cells were allowed to grow in the glass bottom culture dishes for 24 hours prior to the twenty-hour incubation with the dyes. To reduce evaporation due to the extremely small total volume in these experiments, cells which grow at room temperature without a humidified CO<sub>2</sub> incubator were incubated with **1-3** in 96 well plates and then transferred to glass bottom culture dishes for imaging. Imaging was performed with a Zeiss LSM 5 Pascal inverted confocal microscope, equipped with a 40x oil immersion objective lens. Images were captured at 0.88  $\mu$ s/pixel scanning rate. The results for seven cell lines are shown in Table 4.1. Cell lines, ranging from cultured insect and human cancer cells to primary human CD4+ T-cells, were chosen based on availability, relevance to previous studies, and ongoing gene inhibition studies in cell culture.

Under nonfixing conditions, typically two distinct populations of fluorescent cells were observed in a given field. Live cells generally showed relatively dim fluorescence in the cytoplasm, while dead cells generally showed relatively bright fluorescence throughout the cell, including the nucleus. To determine the amount of dead cells in a given sample, solutions of cells on glass bottom culture dishes (with and without pre-incubated **1-3**) were treated with 1  $\mu$ M Sytox Orange for ten minutes and

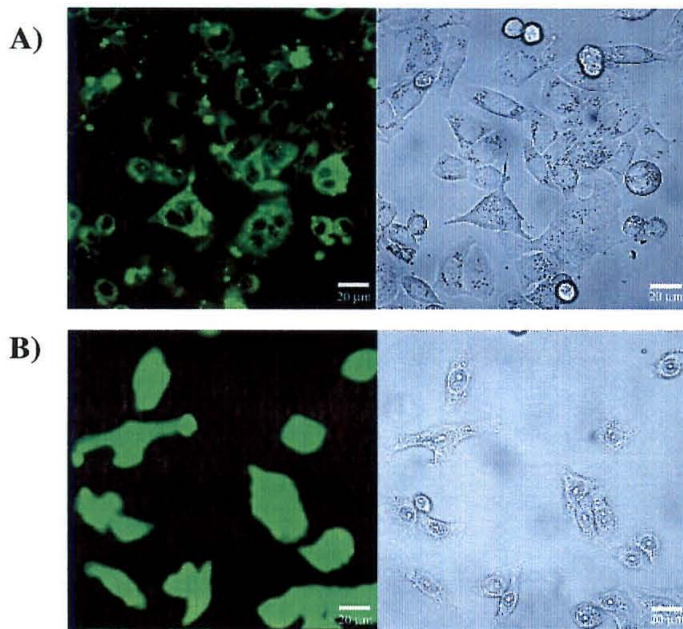


**Table 4.1.** Cellular localization of polyamide-Bodipy conjugates.

Conjugate	SKBR-3 Human breast cancer	CEM human cultured T-cells	T-cells human primary T-cells	NB4 human leukemia	293 human kidney fibroblast	Sf9 Insect	Kc Drosophila
1	cytoplasm	nuclear	nuclear	cytoplasm	cytoplasm	cytoplasm	cytoplasm
2	cytoplasm	nuclear	nuclear	cytoplasm	cytoplasm	cytoplasm	cytoplasm
3	cytoplasm	nuclear	nuclear	cytoplasm	cytoplasm	cytoplasm	cytoplasm

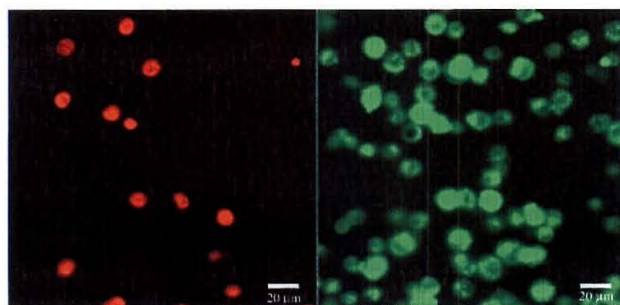
**Table 4.1.** Localization of polyamide-Bodipy conjugates in live cells as determined by confocal microscopy. The designation “nucleus” indicates observation of fluorescence in the interior of the nucleus. The designation “cytoplasm” indicates cellular, non-nuclear fluorescence (see text). Cells were imaged directly following 20-hour incubation with 5  $\mu$ M **1-3** under normal growth conditions for each cell line. SKBR-3 is an adherent, human breast cancer cell line grown in McCoy’s 5A Medium, supplemented with 10% Fetal Bovine Serum and 1% penicillin-streptomycin, at 37° C in 5% CO<sub>2</sub>. CEM is a non-adherent human T-cell line (ATCC, # CCL-119) grown in RPMI 1640 Medium, supplemented with 10% Fetal Bovine Serum and 1% penicillin-streptomycin, at 37° C in 5% CO<sub>2</sub>. Primary human CD4+ T-cells<sup>23</sup> were incubated at 37° C in 5% CO<sub>2</sub>. NB4 is a non-adherent human acute promyelocytic leukemia cell line grown in RPMI 1640 Medium, supplemented with 10% Fetal Bovine Serum and 1% penicillin-streptomycin, at 37° C in 5% CO<sub>2</sub>. 293 is an adherent human kidney fibroblast cell line grown in Dulbecco’s Modified Eagle Medium, supplemented with 10% Fetal Bovine Serum and 1% penicillin-streptomycin, at 37° C in 5% CO<sub>2</sub>. Sf9 is a non-adherent insect (*Spodoptera frugiperda*) cell line grown in TNM-FH insect media, supplemented with 10% Fetal Bovine Serum and 1% penicillin-streptomycin, at 25° C. Kc is a non-adherent *Drosophila* cell line grown in HyQ-CCM3 supplemented with 10% Fetal Bovine Serum, at 25° C.

then imaged directly. As shown in Figure 4.3, the bright nuclear-stained Sf9 cells are dead or damaged as indicated by localization of the membrane impermeant Sytox Orange dead cell stain (Molecular Probes). For each experiment shown in Table 4.1, a control sample of cells was incubated under the same conditions as the sample containing the conjugate. These cells were then imaged to gauge background fluorescence and then treated with Sytox Orange to determine percentage of dead cells in the control sample. Given that cells are imaged directly in growth media, background fluorescence can be significant, but in all cases fluorescence of the dyes and polyamide-dye conjugates far exceeded background. A comparison of the amount of dead cells in the control sample

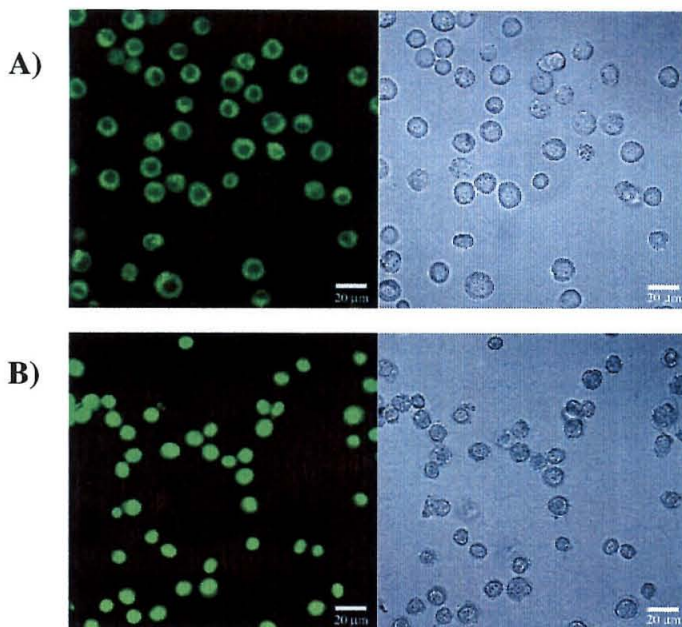
**Figure 4.2**

**Figure 4.2.** A) Conjugate 1 in live SKBR-3 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 μm. B) Conjugate 1 in SKBR-3 cells after the addition of MeOH (1:1 MeOH: media). The fluorescent image is on the left, bright field image on the right. Bar: 20 μm.

and the dye-containing sample allowed the conclusion that 5 μM **1-3** did not cause cell-death under the experimental conditions, but that polyamide-Bodipy conjugates readily localize in the nucleus of dead cells.

**Figure 4.3**

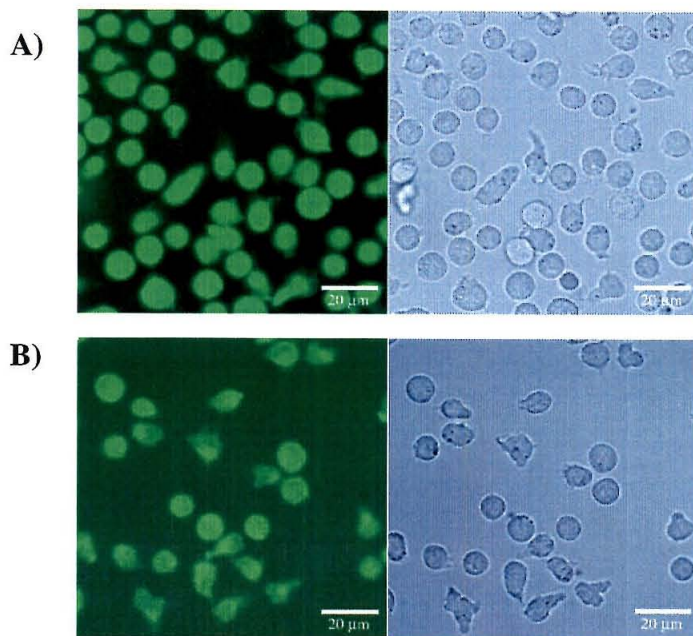
**Figure 4.3.** Conjugate 1 and dead cell stain Sytox Orange in Sf9 cells. The fluorescent emission of Sytox Orange (Rhodamine filter) is shown on the left. The fluorescent emission of conjugate 1 (FITC filter) is shown on the right. Note the different appearance of conjugate 1 in live and dead cells, as distinguished by uptake of Sytox Orange. Bar: 20 μm.

**Figure 4.4**

**Figure 4.4.** A) Conjugate **2** in live Sf9 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 μm. B) Conjugate **2** in Sf9 cells after the addition of MeOH (1:1 MeOH: media). The fluorescent image is on the left, bright field image on the right. Bar: 20 μm.

In controls, free Bodipy FL was either noncellular or localized in the cytoplasm, within the detection limits, in the seven cell lines examined. All live cell lines evaluated were permeable to conjugates **1-3** under the tested conditions (5 μM dye, 20-hour incubation, 25° or 37° C). The designation "cytoplasm" in Table 4.1 is broadly defined as any cellular location other than the interior of the nucleus. The actual location and extent of nonnuclear uptake of **1-3** varied from the punctate pattern observed with **1** in SKBR-3 cells in Figure 4.2A to the more general cytoplasmic staining seen with **2** in Sf9 cells in Figure 4.4A. Rapid nuclear localization was induced in both cases by the addition of methanol (1:1 methanol/media) directly to the solutions of cells and **1** or **2** (Figures 4.2B and 4.4B). The same behavior was observed for other polyamides and cell lines as well as for other common fixing agents, including organic solvents such as ethanol and cross-linking agents formaldehyde and paraformaldehyde.<sup>22</sup>



**Figure 4.5**

**Figure 4.5** A) Conjugate **1** in live CEM cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 μm. B) Conjugate **3** in live CEM cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 μm.

The polyamide-dye conjugates were mainly in the cytoplasm, not the nucleus in five of the cells studied. For the cultured human T-cell (CEM), broad cellular staining including the nucleus was observed with **1** (Figure 4.5A). Fluorescence from conjugate **3** was also observed in the nucleus of live CEM cells, although to a lesser extent (Figure 4.5B). Since conjugates **1** and **3** both contain the same Bodipy dye, the differential nuclear localization may be related to the polyamide structure. Polyamide **3** is an imidazole-rich 2-β-2 hairpin, while polyamide **1** is an eight-ring hairpin containing only an N-terminal imidazole residue. The imidazole rich eight-ring hairpin **2** exhibited a cellular distribution more similar to **1** than to **3**. Thus, the addition of internal beta-alanines (eight ring to 2-β-2) may diminish nuclear localization.

Given the *in vivo* results on HIV-1 inhibition and the observed nuclear localization in the cultured human T-cell line CEM, it was of interest to test the Bodipy

conjugates with primary human CD4+ T-cells. A similar cellular distribution of conjugates **1-3** was observed in primary human CD4+ T-cells as in cultured CEM cells (Table 1).<sup>23</sup> However, it should be noted that over 20 different cell lines have now been examined by the method described here, and nuclear localization of **1-3** was only observed in CEM and primary human T-cells.

## Discussion

While polyamides are larger molecules (MW ~1200) than typical pharmaceuticals, they are considerably smaller than some proteins which can cross through the nuclear pores.<sup>24</sup> Thus, it was believed that if polyamides could gain entry to a cell, they should be able to localize to their DNA targets in the nucleus, given their high affinity for DNA and fast association kinetics.<sup>25</sup> However, while the Bodipy conjugates examined in this study were found to be permeable to all cell lines tested, nuclear localization was only observed in certain T-cells. In this preliminary study the screening conditions chosen (5  $\mu$ M conjugate, 20 hrs incubation) were arbitrary, and the detection limits in this study were lower than under fixed cell conditions. Given that the amount of polyamide in the nucleus necessary to generate a biological effect is unknown, a small amount of conjugate in the nucleus missed by our screen could be significant. As expected, free Bodipy FL was observed to be noncellular or cytoplasmic in all cell lines tested, which indicates that this dye is unlikely to yield false positive results. The observation in CEM and primary human T-cells that the extent of nuclear localization varies with polyamide structure suggests that polyamide redesign might influence cellular distribution. We cannot rule out the possibility that the choice of fluorescent dye may have biased the conjugates against nuclear localization. Further research is in progress to construct other fluorescent polyamides for the purpose of studying cell distribution. The positive results for HIV-1 inhibition in primary T-cells and negative results for inhibition

of HER2/*neu* transcription in SKBR-3 cells, suggest that the Bodipy conjugates reflect the behavior of unmodified polyamides.

Lown and co-workers have recently investigated a fluorescently labeled derivative of distamycin in human ovarian adenocarcinoma (SKOV-3) cells.<sup>26</sup> Distamycin contains three pyrrole carboxamides and is considered the parent DNA binding polyamide molecule.<sup>27</sup> Nuclear localization was not observed within detection limits, and the primary cellular location of the fluorescent derivative was found to be the mitochondria. In conjunction with their studies targeting heterochromatic satellite DNA in *Drosophila*, Laemmli and coworkers have shown specific staining of chromosomes in isolated nuclei with polyamides conjugated with the fluorescent dye Texas Red.<sup>11,12</sup> More recently, these researchers have shown that Texas Red-tandem hairpin conjugates can specifically target both insect and vertebrate telomere repeat sequences, and can be used for rapid estimations of relative telomere length *in vitro*.<sup>28</sup> Extending these studies to intact cells, rapid uptake, nuclear localization, and staining of telomeres was observed in Sf9 cells. However, these experiments were performed under fixing conditions with methanol.<sup>28</sup>

There is no chemical reason for the dramatic effect of fixing agents to be specific for polyamide-Bodipy conjugates. It is possible that any molecule that has a high-affinity receptor (i.e., DNA) from which it is excluded by cellular processes would exhibit the same behavior. However, in a very recent study with fluorescein-labeled oligonucleotides, greater nuclear localization was in fact observed in live cells as compared to fixed cells.<sup>29</sup> In surveying the literature, many published reports on cellular uptake and localization of designed ligands show data from fixed cells, without evidence of having performed a live cell control. These results suggest that such studies should demonstrate localization in live cells.

The finding that polyamide-Bodipy conjugates localize to the nucleus of CEM and primary human T-cells is significant. This result is consistent with the previous HIV-

1 inhibition study,<sup>10</sup> and provides impetus for further screening to find other cell lines amenable to nuclear localization. The mechanisms by which polyamide-Bodipy conjugates enter live cells and are excluded from the nucleus are unknown at this point. These mechanisms could be distinct and also could be variable between different cell lines. Understanding the mechanisms of cellular entry and localization is an important issue for reengineering second-generation polyamides with specified nuclear uptake properties.

## Conclusions

In the majority of live cells tested, polyamide-Bodipy conjugates were observed to localize mainly in the cytoplasm, not the nucleus. These results suggest that nuclear localization is likely a hurdle for the development of polyamides as *in vivo* modulators of gene transcription. We are currently in the process of developing modified polyamides with enhanced nuclear localization properties which will be reported in due course.

## Acknowledgements

We are grateful to Scott Fraser and Hayong Lim for use of the Caltech Biological Imaging Facility and assistance in confocal microscopy. Special thanks to Inder Nangiana of the Caltech Protein Expression Facility for help with cell culture. We are indebted to Aseem Z. Ansari, Joel Pomerantz, Joel Gottesfeld, and Carl Parker for providing cells and helpful suggestions. We are grateful to the National Institutes of Health (GM 57148) for research support, the Ralph M. Parsons Foundation for a predoctoral fellowship to J.M.B., and the American Cancer Society for a postdoctoral fellowship to P.S.A.

## References and Notes

- (1) Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688-693.
- (2) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.
- (3) Gottesfeld, J. M.; Turner, J. M.; Dervan, P. B. *Gene Expression* **2000**, *9*, 77-91.
- (4) Chang, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 4856-4864.
- (5) Wurtz, N. R.; Dervan, P. B. *Chem. Biol.* **2000**, *7*, 153-161.
- (6) Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2001**, *123*, 8657-8661.
- (7) Mapp, A. K.; Anasari, A. Z.; Ptashne, M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3930-3935.
- (8) Anasari, A. Z.; Mapp, A. K.; Nguyen, D. H.; Dervan, P. B.; Ptashne, M. *Chem. Biol.* **2001**, *8*, 583-592.
- (9) Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. *J. Mol. Biol.* **2001**, *309*, 615-629.
- (10) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895.
- (11) Janssen, S.; Durussel, T.; Laemmli, U. K. *Mol. Cell* **2000**, *6*, 999-1011.
- (12) Janssen, S.; Cuvier, O.; Muller, M.; Laemmli, U. K. *Mol. Cell* **2000**, *6*, 1013-1024.
- (13) Slamon, D. J.; Clark, G. M.; Wong, S. G.; Levin, W. J.; Ullrich, A.; McGuire, W. L. *Science* **1987**, *235*, 177-182.
- (14) Singleton, T. P.; Strickler, J. G. *Pathol. Annu.* **1992**, *27*, 165-190.
- (15) Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A. *J. Biol. Chem.* **2000**, *275*, 24246-24254.
- (16) Dervan, P. B.; Gottesfeld, J. M.; Beerman, T. A. Unpublished results.
- (17) Karolin, J.; Johansson, L. B. A.; Strandberg, L.; Ny, T. *J. Am. Chem. Soc.* **1994**, *116*, 7801.



- (18) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
- (19) Mapp, A. K.; Dervan, P. B. *Tetrahedron Lett.* **2000**, *41*, 9451-9454.
- (20) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.
- (21) Fixing procedure: cover slips were washed 3X in ice-cold PBS, followed by 5-minute incubation in ice-cold methanol. Cover slips were rehydrated/washed 3X, 5 minutes each with ice-cold PBS, rinsed in ddH<sub>2</sub>O and mounted.
- (22) Leslie, S. J. Ph.D. Dissertation, Rosewell Park Cancer Institute: Buffalo, NY, 2001.
- (23) We are grateful to Dr. Gaston Picchio (Mosier lab, Scripps Research Institute) for the acquisition and purification of primary human CD4+ T-cells from healthy donors.
- (24) Rout, M. P.; Aitchison, J. D. *J. Biol. Chem.* **2001**, *276*, 16593-16596.
- (25) Baliga, R.; Baird, E. E.; Herman, D. M.; Melander, C.; Dervan, P. B.; Crothers, D. M. *Biochemistry* **2001**, *40*, 3-8.
- (26) Sharma, S. K.; Morrissey, A. T.; Miller, G. G.; Gmeiner, W. H.; Lown, J. W. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 769-772.
- (27) Bailly, C.; Chaires, J. B. *Bioconjugate Chem.* **1998**, *9*, 513-538.
- (28) Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218-3228.
- (29) Rapozzi, V.; Cogoi, S.; Spessotto, P.; Risso, A.; Bonora, G. M.; Quadrifoglio, F.; Xodo, L. E. *Biochemistry* **2002**, *41*, 502-510.

## Chapter 4B:

### Enhanced Nuclear Localization of Polyamide-Dye Conjugates<sup>1</sup>

#### Overview

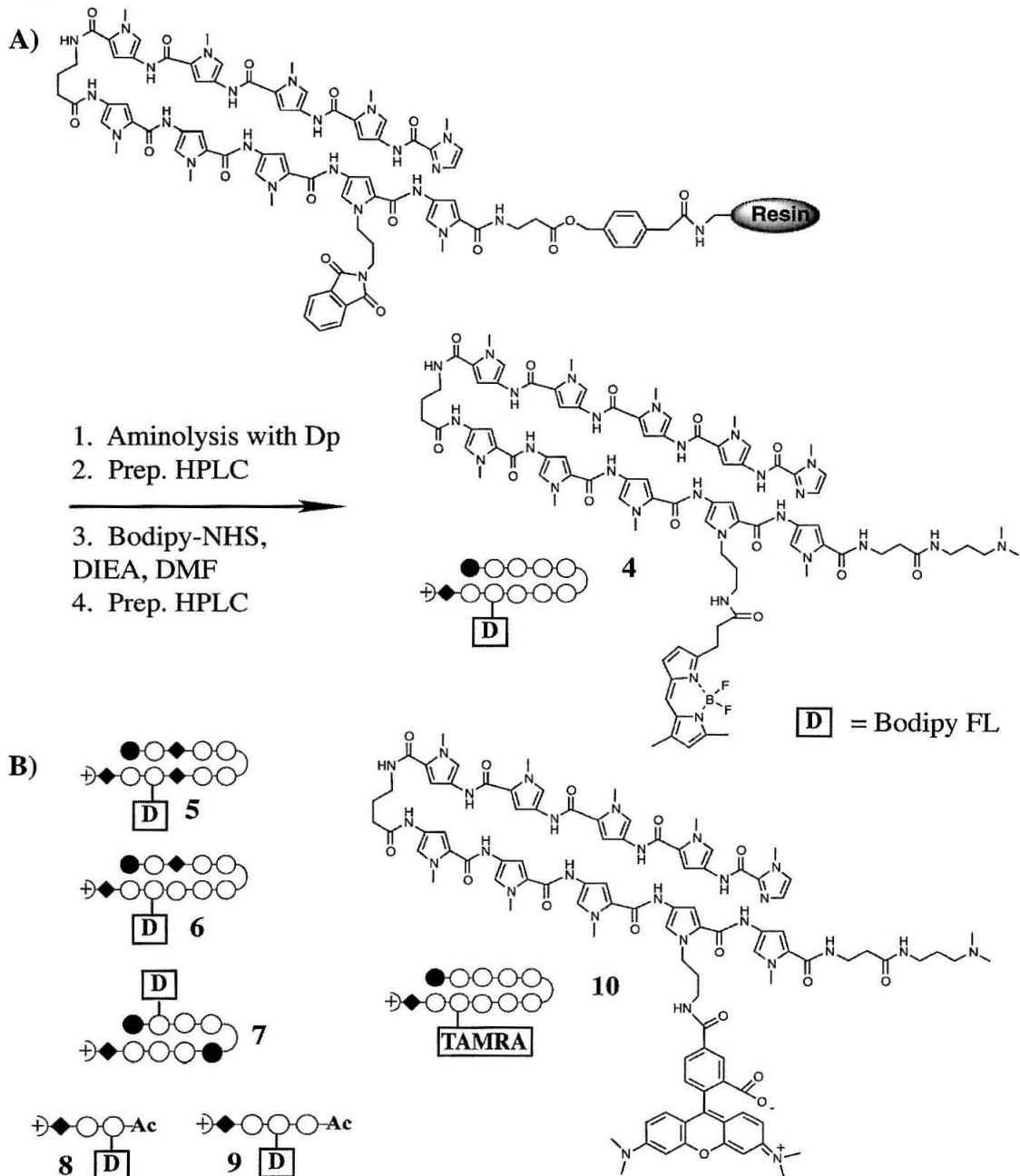
As described in Chapter 4A, screening polyamide-Bodipy conjugates in over twenty cell lines uncovered only two cell lines (primary human CD<sup>+</sup> T-cell and cultured human T-cell, CEM) in which detectable levels of nuclear localization were observed in live cells. This chapter describes efforts to enhance nuclear localization in the majority of cell lines. It is noted explicitly that the results described here apply to polyamide-dye conjugates, in almost all cases conjugates of Bodipy FL. Independent assays without the fluorescent tag will be necessary to confirm that these results apply in the absence of the dye. It is also possible that the measures taken will be more effective in the absence of the dye. With these caveats, the imaging experiments described here provide a consistent, though in some cases confusing, set of data for conjugate uptake and localization. All imaging experiments were performed as described in Chapter 4A for live cells, using the standard conditions of 5  $\mu$ M conjugate and ~20 hour incubation, unless otherwise noted. The results will be used to determine which classes of compounds are best suited for functional assays in particular cell lines. Multiple solutions for several cell lines have emerged, and on-going experiments should define further solutions.

Characterization of standard polyamide-fluorescent dyes is presented first, followed by conjugates of different polyamide motifs. In particular these studies suggest the importance of molecular weight for nuclear localization in some, but not all, cell lines examined. Next conjugates of the alkylating agent chlorambucil<sup>2</sup> are described, followed by small molecule conjugates for receptor-mediated uptake.<sup>3-5</sup> The best results in terms of generality of nuclear localization across a wide variety of cell lines have come from carrier peptides,<sup>6</sup> such as the well known TAT<sup>7</sup> and the newer derivative R9.<sup>8</sup> However,

the peptides which were successful were also highly positively charged. The effects on DNA binding of polyamide-cationic peptide conjugates are discussed. The last portion of this chapter details attempts to find a balance between enhanced nuclear localization and the specificity that is the hallmark of DNA binding polyamides.

## Results and Discussion

**Polyamide-Dye Conjugates.** In addition to polyamide-Bodipy conjugates **1-3**, compounds **4-10** were also synthesized by standard methods (Figure 4.6). Conjugates **4** and **5** are analogs of polyamides which prevent binding of the transcription factor ESX to the HER2 promoter.<sup>9</sup> Compound **4** was the first Bodipy conjugate used to address cell permeability of polyamides that were active HER2 transcription inhibitors *in vitro*, but not in breast cancer cells (see Chapter 3). Interestingly, the localization of conjugates **4** and **5**, which only differ by the central pairing (Py/Py or  $\beta/\beta$ ), varied when the imaging experiments were performed under fixing conditions. Conjugate **4** strongly stained the nucleus of SKBR-3 while conjugate **5** appeared to localize at or near the nuclear membrane.<sup>10</sup> The discovery of the effect of organic solvents on polyamide-dye conjugates was made in control experiments probing the differences between **4** and **5**. These differences are likely not physiologically relevant as in live SKBR-3 cells neither conjugate localizes to the nucleus or the nuclear membrane, instead they show a punctate pattern in the cytoplasm (also observed for **1** in SKBR-3 cells, Figure 4.2A). Conjugates **4** and **5** have been studied in a variety of other cell lines as well, although their utility under the live-cell experimental conditions described in Chapter 4A is limited by a tendency to precipitate on the glass bottom culture dishes due to poor solubility. Conjugate **6**, with a central  $\beta$ /Py pairing was originally synthesized to probe the different fixed-cell localization between **4** and **5**, and has been investigated sparingly in live (or fixed) SKBR-3 cells. Conjugate **7** has also thus far only been investigated in SKBR-3 cells, although its solubility properties are similar to **1-3** and it could be of general utility.

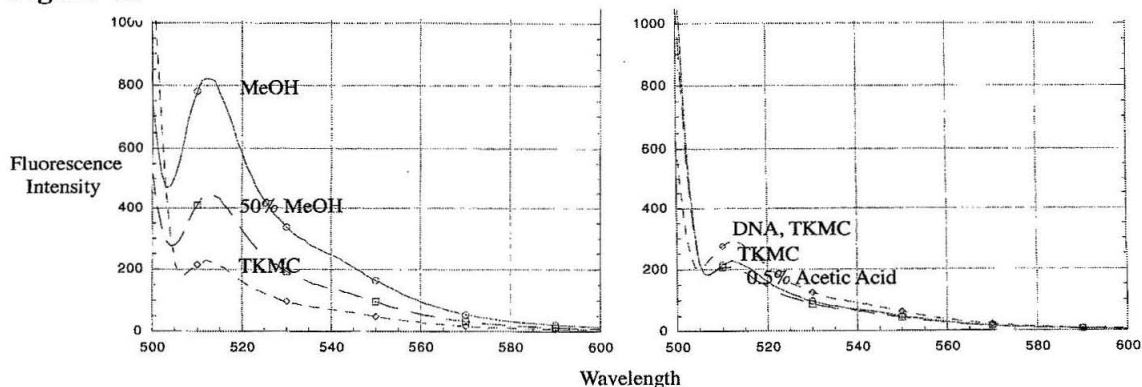
**Figure 4.6**

**Figure 4.6. A)** Representative synthesis of a polyamide-Bodipy conjugate (**4**) from a standard polyamide-PAM resin prepared by solid-phase synthesis. **B)** Polyamide-dye conjugates shown in ball-and-stick format, with Im as a filled circle, Py as an open circle,  $\beta$ -alanine ( $\beta$ ) as a grey diamond,  $\gamma$ -aminobutyric acid as a curved lined, and dimethylaminopropylamine (Dp) as a plus sign. The chemical structure of **10**, which is identical to **4** with tetramethylrhodamine in place of Bodipy is also shown.

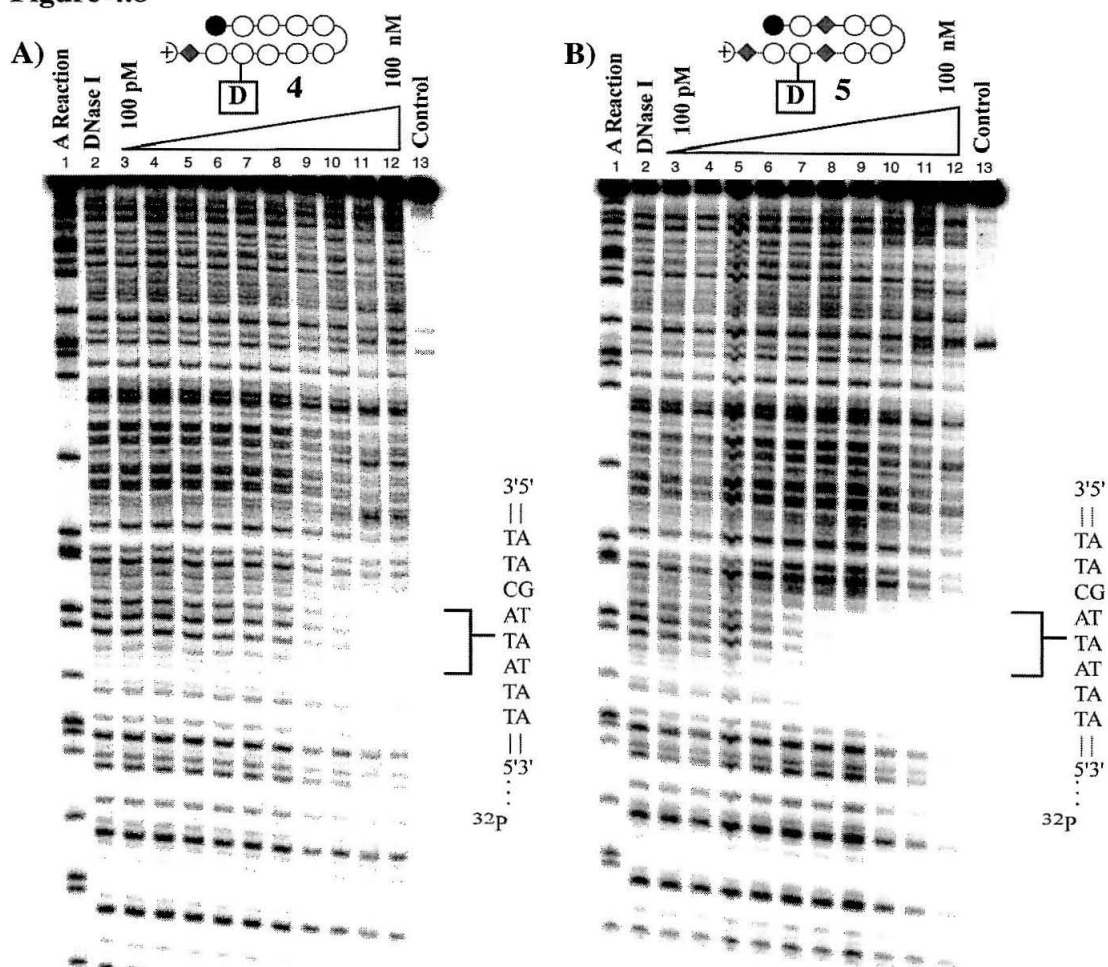
Like **3**, the polyamide portion of **7** is derived from an *in vivo* inhibitor of HIV-1 replication.<sup>11</sup> Compounds **8** and **9** with only two and three Py rings, respectively, are netropsin and distamycin-like Bodipy conjugates. Both were cytoplasmic in SKBR-3 cells, although interestingly, significantly more of the two-ring conjugate **8** was inside the cells than the three-ring conjugate **9**. These smaller conjugates are currently being investigated in a variety of cell lines as part of a program to ascertain the role of molecular weight in the nuclear localization of polyamide-Bodipy conjugates.

Compound **10** was the first non-Bodipy polyamide-dye conjugate tested for uptake and localization by the Dervan and Beerman laboratories. The tetramethyl-rhodamine (TAMRA) conjugate has the same parent polyamide (ESX-binding inhibitor on the HER2 promoter) as **4**. Unfortunately, it is even less soluble than **4**, and examination of this dye in live cells on the glass bottom culture dishes is hampered by precipitation. Subsequently several other hairpin conjugates with dyes such as TAMRA, Cy3, and Cy5 have been investigated.<sup>12</sup> Thus far, the results of confocal microscopy screening of various cell lines with the non-Bodipy conjugates such as **10** parallel the results for Bodipy conjugates. This suggests that Bodipy in of itself is not directing the

**Figure 4.7**



**Figure 4.7.** Fluorescence of Bodipy-polyamide conjugate **4** under various conditions. TKMC is a buffer containing 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>, commonly used for DNase I footprinting experiments. DNA refers to an equimolar quantity of the match duplex 5'-GGGAAGTAT-AAGGG-3', derived from the binding site of **4** on the HER2 promoter (Figure 4.8). 0.5% acetic acid is representative of the pH of acidic cellular vesicles such as lysosomes.

**Figure 4.8**

**Figure 4.8.** Quantitative DNase I footprint titration experiments on a 5'-<sup>32</sup>P-labeled 188-bp DNA fragment of the HER2 promoter.<sup>9</sup> All reactions contain 15 kcpm DNA fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. **A)** Lane 1, A reaction; lane 2, DNase I standard; lanes 3-12, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide-Bodipy conjugate 4, respectively; lane 13, intact DNA. **B)** Lane 1, A reaction; lane 2, DNase I standard; lanes 3-12, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide-Bodipy conjugate 5, respectively; lane 13, intact DNA.

localization of the polyamides, although in all these cases a significant amount of molecular weight has been added to the polyamide. Efforts are underway to determine if a fluorescent moiety can be incorporated into the polyamide backbone, without significantly altering the DNA binding properties.

It was noted in Chapter 4A that the fluorescence of the Bodipy-conjugates was unaffected by the presence of cognate DNA. This appears to be the exception, rather than the rule, for polyamide-dye conjugates. With fluorophores such as TAMRA, Cy3, Cy5, and Oregon Green, polyamide-dye conjugates have shown the exciting property of enhanced fluorescence when bound to match DNA sequences.<sup>13</sup> TAMRA conjugate **10** shows a 15-fold fluorescence enhancement in the presence of match DNA. In contrast Bodipy conjugate **4** shows less than 1.5-fold enhancement with the same match duplex, considerably less than the 4-fold difference between 100% MeOH and buffer (Figure 4.7). The mechanism of fluorescence enhancement for the non-Bodipy fluorescent conjugates is currently under investigation. The fluorescence enhancement exhibited by **10** has been useful for investigations in the Beerman laboratory aimed at increasing nuclear localization of polyamides with external reagents and processes, such as liposomes and co-transfection with DNA.

**DNA Binding of Polyamide-Dye Conjugates.** The affinity and specificity of representative Bodipy conjugates **1,4-7** and TAMRA conjugate **10** have been investigated by quantitative DNase I footprinting titration experiments (Table 4.2).<sup>14</sup> Representative gels for conjugates **4** and **5** are shown in Figure 4.8. In general the conjugates were quite specific for their match sites, although some coating of the DNA (non-specific binding to all sequences) was noted for **6** and **10** at high concentrations. The affinities of the eight- and ten-ring conjugates were 75- to 100-fold lower than the parent polyamides, while affinity of the 2- $\beta$ -2 conjugate **5** is 9-fold less than the parent (doubly charged) compound ( $K_a = 8.7 \times 10^9 \text{ M}^{-1}$ ). The affinity of **5** ( $K_a = 1 \times 10^9 \text{ M}^{-1}$ ) is rather high for a polyamide-dye conjugate,<sup>12</sup> suggesting that the placement of a  $\beta$ -alanine or specifically a  $\beta/\beta$  pair adjacent to a modified Py ring may be beneficial for the affinity of similarly modified conjugates. Compound **6**, which has a  $\beta$ -alanine in the N-terminal strand, but a Py

**Table 4.2.** Equilibrium association constants of polyamide-dye conjugates.

Polyamide	Binding Site	$K_a$	DNA Fragment
<b>4</b>	5'-gAAGTATAag-3'	$1.4 \times 10^8 \text{ M}^{-1}$	HER2 Promoter
<b>5</b>	5'-gAAGTATAag-3'	$1.0 \times 10^9 \text{ M}^{-1}$	HER2 Promoter
<b>6</b>	5'-gAAGTATAag-3'	$6.2 \times 10^7 \text{ M}^{-1}$	HER2 Promoter
<b>10</b>	5'-gAAGTATAag-3'	$3.6 \times 10^7 \text{ M}^{-1}$	HER2 Promoter
<b>1</b>	5'-tTAGTATTtg-3'	$5.0 \times 10^7 \text{ M}^{-1}$	pJT8
<b>7</b>	5'-tTAGTACTtg-3'	$3.3 \times 10^8 \text{ M}^{-1}$	pJT8

**Table 4.2.** Equilibrium association constants ( $\text{M}^{-1}$ ) for polyamide-Bodipy conjugates **4-6** and TAMRA conjugate **10** with a 188-bp DNA fragment of the HER2 promoter,<sup>9</sup> and polyamides **1** and **7** with restriction fragment pJT8.<sup>15</sup> Values reported are the mean values obtained from at least three DNase I footprint titration experiments. Assays were performed at 22° C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

adjacent to the modified position, showed considerably lower affinity ( $K_a = 6.2 \times 10^7 \text{ M}^{-1}$ ). The high specificity of many of these compounds and the equilibrium association constants above  $10^7 \text{ M}^{-1}$  suggest that the Bodipy conjugates are good models of standard hairpin polyamides.

**Confocal Microscopy.** A summary of cell lines not described in Chapter 4A that have been examined with at least one of polyamide-dyes **1-5** is given in Table 4.3. In the last two entries on the table, a pre-B-cell (103Bcl2) and a plasmacytoma cell (S194), nuclear localization was observed with standard hairpin-dyes, in very recent experiments. The Dervan laboratory will be actively pursuing functional assays in these cell lines. In the other cell lines, the Bodipy conjugates were concentrated either broadly throughout the cytoplasm, as in the insect cell line Hi-5 (Figure 4.9A), or in a more localized, punctate pattern, as in the human prostate cancer cell line LnCap (Figure 4.9B).

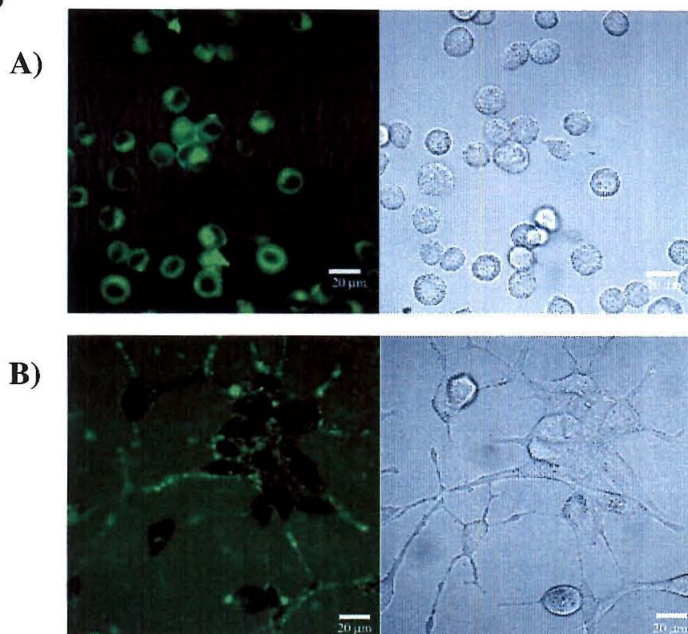


**Table 4.3.** Screening of standard hairpin-Bodipy conjugates **1-5**.

Cell line	Nuclear Localization	Cell line	Nuclear Localization
MCF-7 <i>breast cancer</i>	—	Mel <i>erythroid cancer</i>	—
T47D <i>breast cancer</i>	—	PC-3 <i>prostate cancer</i>	—
MOATB3468 <i>breast cancer</i>	—	LnCap <i>prostate cancer</i>	—
Hi-5 <i>insect</i>	—	786-O <i>renal carcinoma</i>	—
S2 <i>Drosophila</i>	—	A-498 <i>renal carcinoma</i>	—
He-La <i>cervical cancer</i>	—	769-P <i>renal carcinoma</i>	—
Jurkat <i>T-cell</i>	—	HTB-44 <i>renal carcinoma</i>	—
K-562 <i>bcr/abl</i>	—	103Bcl2 <i>pre-B-cell</i>	+
A6 <i>Xenopus kidney fibroblast</i>	—	S194 <i>plasmacytoma</i>	+

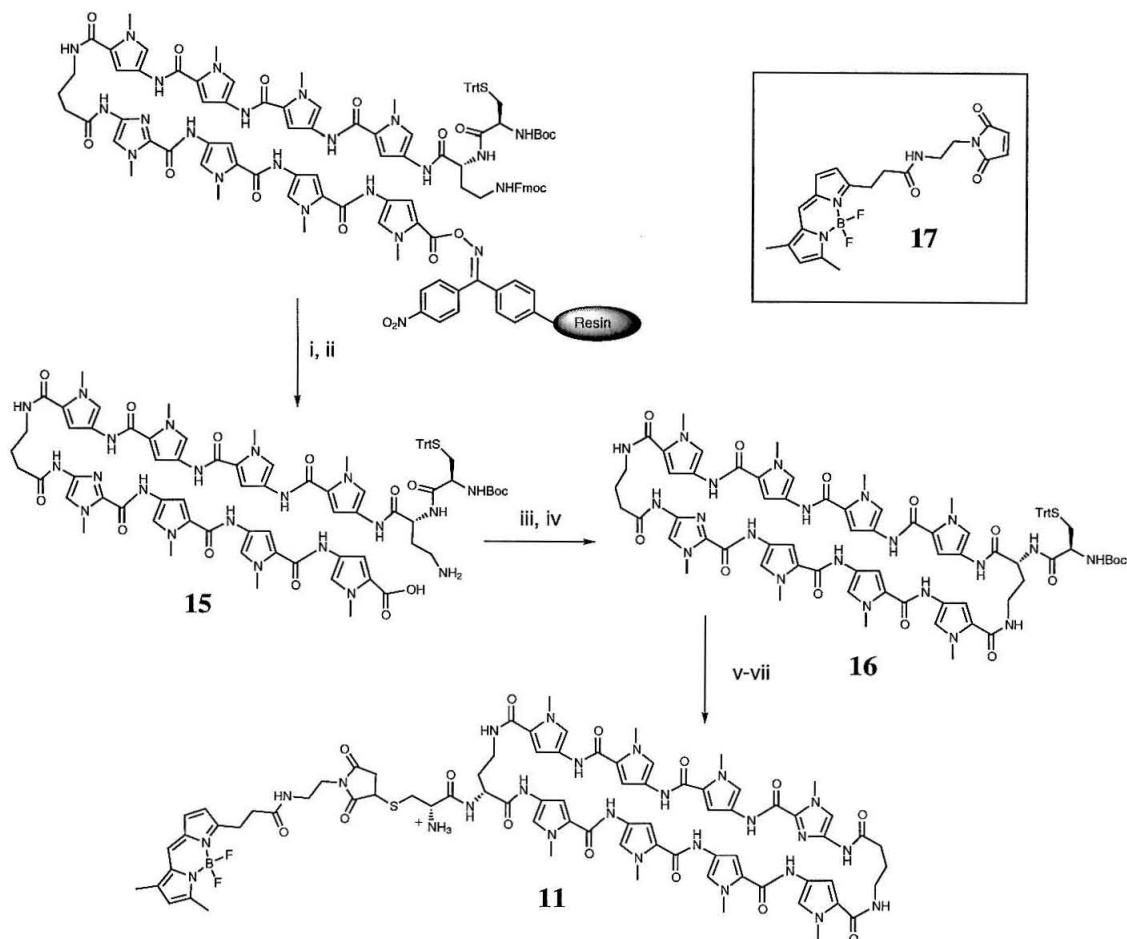
**Table 4.3.** Additional cell lines where at least one of **1-5** has been screened (usually **1-3**) according to the protocol described in Chapter 4A. Cell lines are human-derived except where noted. Cells were imaged directly following 20-hour incubation with 5  $\mu$ M conjugate under normal growth conditions for each cell line.

**Bodipy conjugates of different polyamide motifs.** Since hairpin conjugates did localize to the nucleus of the majority of live cells tested, it was of interest to investigate other polyamide motifs. Bodipy conjugates of a cycle<sup>16</sup> (**11**), a tandem hairpin dimer<sup>17</sup> (**12**), and two one-to-one binding polyamides<sup>18</sup> (**13-14**) were examined in a variety of cell lines (Figures 4.10-4.12, Table 4.4). Whereas hairpin conjugates such as **1-5** are readily accessible, synthesis of cyclic polyamides is known to be challenging. In this case newly developed protocols for polyamide synthesis on oxime resin (Chapter 5A) were utilized to obtain the C-terminal acid intermediate **15** (Figure 4.10), after extensive attempts to

**Figure 4.9**

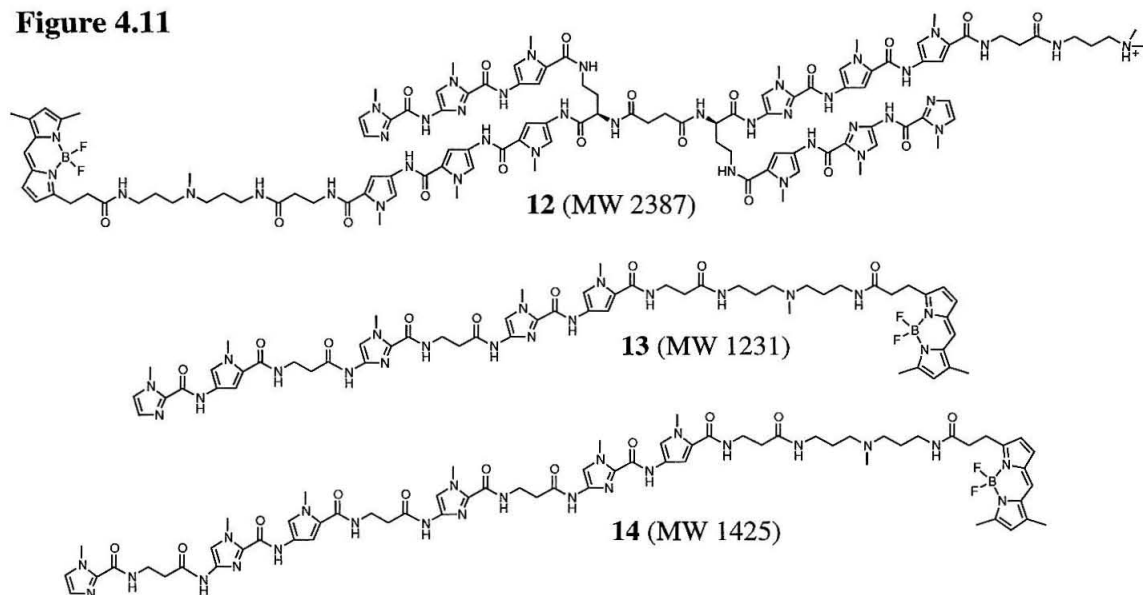
**Figure 4.9.** A) Conjugate **3** in live Hi-5 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20  $\mu\text{m}$ . B) Conjugate **1** in live LnCap cells. The fluorescent image is on the left, bright field image on the right. Bar: 20  $\mu\text{m}$ .

cyclize eight-ring polyamides directly from the oxime resin proved to be unsuccessful. While the DBU-assisted cleavage from the resin to generate **15** was a lower-yielding reaction than the corresponding cleavage without the trityl-protected cysteine residue, the diphenylphosphorylazide (DPPA)-mediated cyclization<sup>16</sup> actually occurred in higher yield (a quite respectable 30% recovery following preparative HPLC purification) with this bulky substituent close to the nucleophilic amine than without it (8% recovery). The Bodipy maleimide unit **17** was reacted with the free thiol following cleavage of the trityl and Boc protecting groups from cysteine-appended cyclic polyamide **16**, to complete the Bodipy-cyclic polyamide conjugate **11**. It was envisioned that the cycle might behave quite differently from the hairpin-dyes, because of its constrained structure. However, thus far the cycle conjugate had similar uptake and localization properties to the hairpin-dyes, in that it was cell-permeable to the four cell lines tested, but only localized to the nucleus in live CEM cells.

**Figure 4.10**

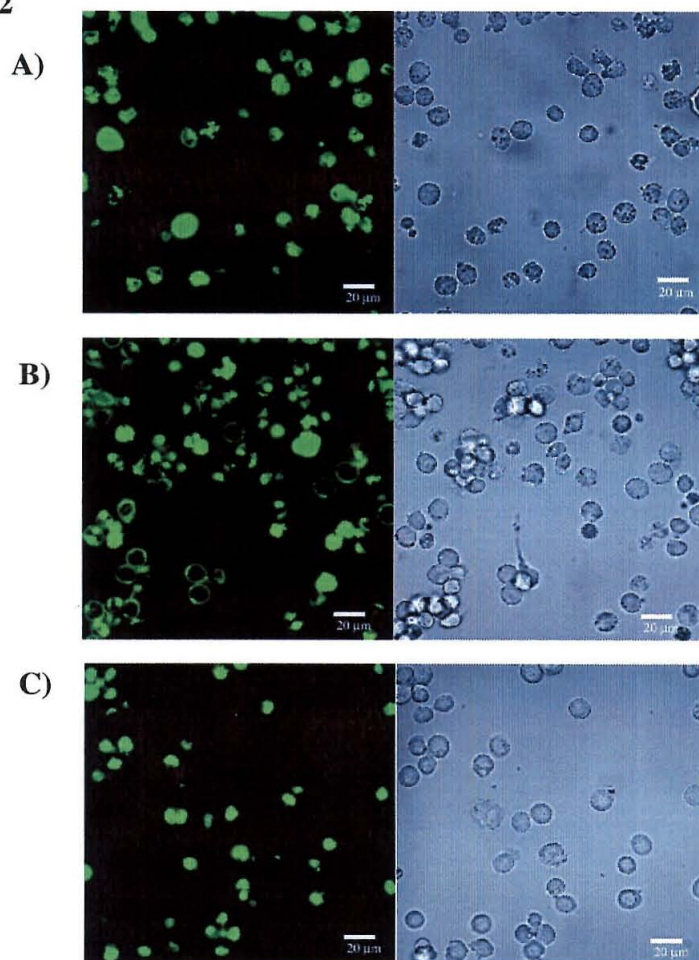
**Figure 4.10.** Synthesis of the Bodipy-cyclic polyamide conjugate **11**: (i)  $\text{H}_2\text{O}$ , DMF, DBU,  $37^\circ\text{C}$ , 72 hrs; (ii) preparative HPLC purification; (iii) DPPA,  $\text{K}_2\text{CO}_3$ , DMF, 4.5 hrs; (iv) preparative HPLC purification; (v) 33% TFA/ $\text{CH}_2\text{Cl}_2$ ,  $\text{Et}_3\text{SiH}$ , 15 min.; (vi) 100 mM  $\text{Na}_2\text{HPO}_4$ ,  $\text{H}_2\text{O}$  (pH = 7), DMF, **17**, TCEP, 30 min. (vii) preparative HPLC purification.

The tandem and one-to-one binding polyamides are on opposite sides of the polyamide molecular weight spectrum. Given the high molecular weight, it was surprising that the Bodipy-tandem conjugate **12** was not only cell-permeable but also localized to the nucleus of live CEM cells. However, like the cycle, it was not observed in the nucleus of other live cells, and appeared to be somewhat less permeable to the cytoplasm than the standard hairpin-dyes in some cases. The one-to-one binder conjugates **13** and **14** have only recently been investigated, however, they have already

**Figure 4.11****Figure 4.11.** Chemical structures of tandem hairpin dimer and one-to-one binding Bodipy conjugates.**Table 4.4.** Bodipy conjugates of different polyamide motifs.

	11	12	13	14
<b>CEM</b>	nuclear	nuclear	nuclear	cytoplasm
<b>NB4</b>	cytoplasm	cytoplasm	nuclear	cytoplasm
<b>SKBR-3</b>	-	cytoplasm	cytoplasm	cytoplasm
<b>MCF-7</b>	-	cytoplasm	-	-
<b>293</b>	-	cytoplasm	cytoplasm	cytoplasm
<b>Sf9</b>	-	cytoplasm	nuclear	nuclear
<b>Kc</b>	-	cytoplasm	nuclear	nuclear
<b>Hi5</b>	-	cytoplasm	-	-
<b>Mel</b>	-	-	cytoplasm	cytoplasm
<b>LnCap</b>	cytoplasm	-	-	-
<b>PC-3</b>	cytoplasm	-	-	-

**Table 4.4.** Localization of polyamide-Bodipy conjugates **11-14** in live cells as determined by confocal microscopy. The designation “nucleus” indicates observation of fluorescence in the interior of the nucleus. The designation “cytoplasm” indicates cellular, nonnuclear fluorescence. The designation “-” indicates that the experiment has not yet been performed. Cells were imaged directly following 20-hour incubation with 5  $\mu$ M **11-14** under normal growth conditions for each cell line.

**Figure 4.12**

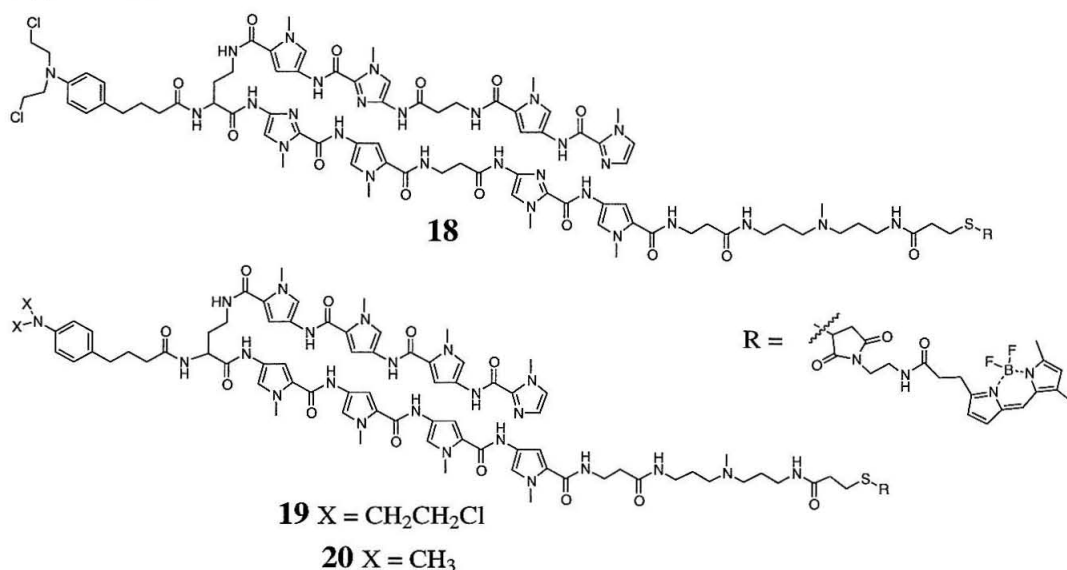
**Figure 4.12.** A) Conjugate **13** in live NB4 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 µm. B) Conjugate **14** in live NB4 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 µm. C) Conjugate **13** in live Sf9 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 µm.

provided very exciting results. While not a completely general solution, nuclear localization of **13** was observed in four cell lines, including CEM, the human leukemia cell line NB4 (Figure 4.12A), and the insect cell line Sf9, where the conjugate appeared to localize exclusively to the nucleus (Figure 4.12C). Conjugate **14** also localized strongly to the nucleus in Sf9 cells, and both one-to-one binder conjugates were observed in the nucleus of *Drosophila* Kc cells. While a whole organism is tremendously different from one of its cells, this result is at least consistent with the exciting work of Laemmli

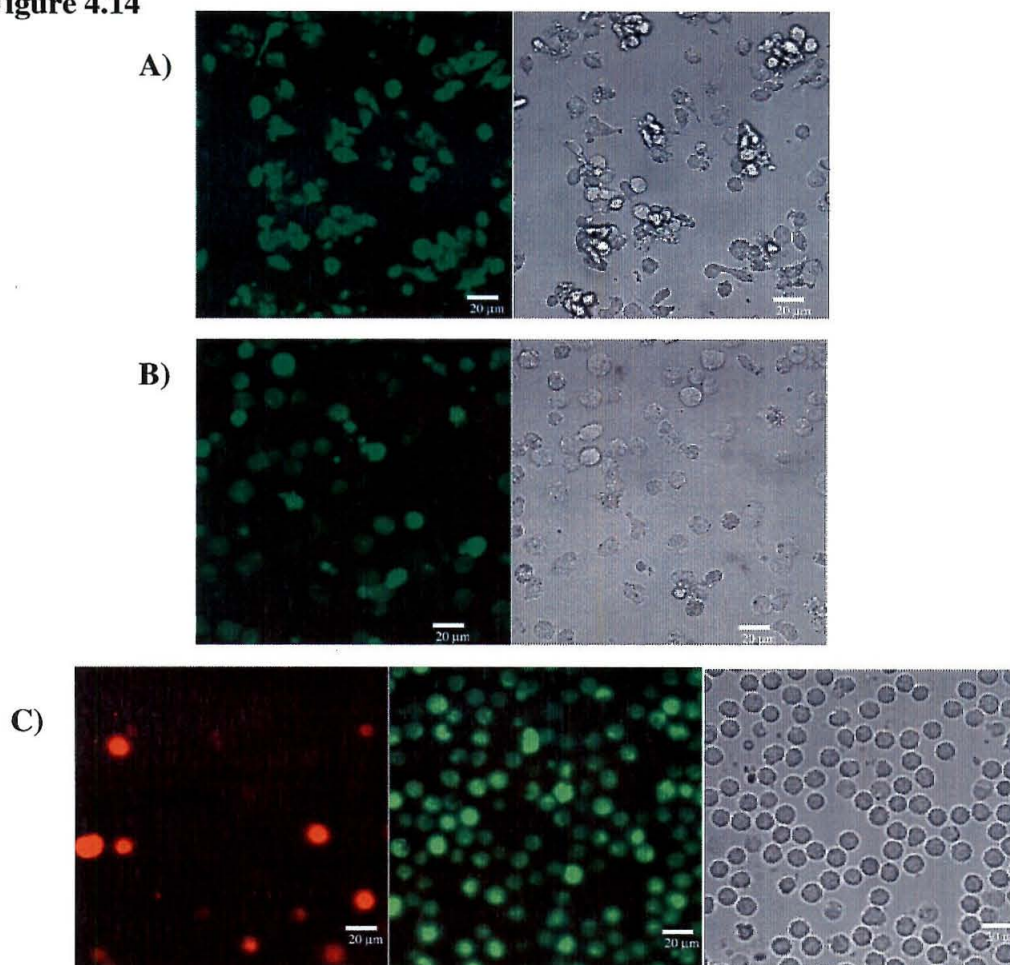


and co-workers, who showed that similar one-to-one binding polyamides caused phenotypic changes in live *Drosophila* as a result of interaction with specific satellite DNA repeat sequences.<sup>19,20</sup> Compared to the standard hairpin-dyes, the results of **13** and **14** suggest that molecular weight is likely a factor for nuclear localization in insect cells such as Kc and Sf9. In human cells the importance of molecular weight is less clear, although there appears to be some dependence, as the slightly larger **14** did not localize to the nucleus of NB4 cells (Figure 4.13B). However, **14** was also only observed in the cytoplasm of CEM cells, where conjugates as large as the tandem dimer **12** were observed in the nucleus. Molecular weight is usually mentioned as an issue in cell permeability, its role once a compound is inside of cells is less clear. Nevertheless, these recent results represent an important lead, as it is certainly possible that in other human cell lines conjugates smaller than **13** would display enhanced nuclear localization. As previously mentioned, a study examining a series of progressively lower molecular weight polyamide-Bodipy conjugates has been initiated. These results also raise the issue that the molecular weight of Bodipy may contribute to the localization properties of the conjugates, as it is worth noting that the molecular weight of conjugate **15** is very similar to the molecular weight of a standard unmodified eight-ring polyamide. This highlights the importance of efforts to develop intrinsically fluorescent polyamides and other methods to gauge whether unmodified polyamides can reach their DNA targets in live cells.

**Chlorambucil-polyamide-Bodipy conjugates.** The DNA-alkylating agent chlorambucil (CHL) has been attached to hairpin polyamides to produce sequence-specific alkylating agents. The laboratories of both Terry Beerman and Joel Gottesfeld have observed effects such as toxicity and DNA-alkylation in a variety of cell lines.<sup>21</sup> Thus, it was interesting to test polyamide-Bodipy-CHL conjugates for nuclear localization by confocal microscopy. Three conjugates were examined in a variety of

**Figure 4.13****Figure 4.13.** Structures of CHL-polyamide-Bodipy conjugates **18-19** and non-alkylating analog **20**.

cell lines (Figures 4.13 and 4.14, Table 4.5). In the majority of the cell lines tested, the alkylator conjugates **18** and **19** did not localize to the nucleus or appear to cause appreciable cell death. The alkylator conjugates did localize to the nucleus of three human cell lines: NB4 (Figure 4.15A and 4.15B), CEM, and significantly the primary CD4<sup>+</sup> T-cells (Figure 4.15C). However, even with nuclear localization, conjugates **18** and **19** did not cause cell death, as measured by costaining with sytox orange. (Cell growth inhibition/cell cycle arrest were not explicitly screened for. Qualitatively, a diminution in the amount of cells at the 20-hour imaging point was not observed.) Addition of the dye will likely decrease the DNA-binding affinity of the conjugate, which will lower alkylation potential, although this alone may account for differences in toxicity. Particularly in the cells where the conjugate was not observed in the nucleus, Bodipy may alter its localization. Alternatively, the conjugates may faithfully reflect the behavior of CHL-polyamides, but, without attenuation by attachment of a dye, a small amount reaching the nucleus (below the detection limit in the confocal microscopy experiments) may be enough to cause extensive toxic effects. It is also possible that

**Figure 4.14**

**Figure 4.14.** **A)** Eight-ring hairpin-CHL-Bodipy conjugate **19** in live NB4 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 μm. **B)** 2-β-2-CHL-Bodipy conjugate **18** in live NB4 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 μm. **C)** 2-β-2-CHL-Bodipy conjugate **18** and dead cell stain Sytox Orange in live primary human CD4+ T-cells. The fluorescent emission of Sytox Orange (Rhodamine filter) is shown on the left. The fluorescent emission of conjugate **1** (FITC filter) is shown in the center. The bright field image on the right. Bar: 20 μm.

alkylation of proteins is responsible for toxicity, and that this is also attenuated by attachment of the dye. Alkylation of DNA or perhaps proteins is likely a factor in NB4 cells, since **20** a non-alkylating analog of **19** did not localize to the nucleus, although specific physical properties of the extra  $-\text{CH}_2\text{CL}$  groups of **19** not related to alkylation have not been ruled out. Nuclear localization of the non-alkylating CHL hydrolysis product ( $x = \text{CH}_2\text{CH}_2\text{OH}$ ) is also a possibility. The polyamide portion of the molecule



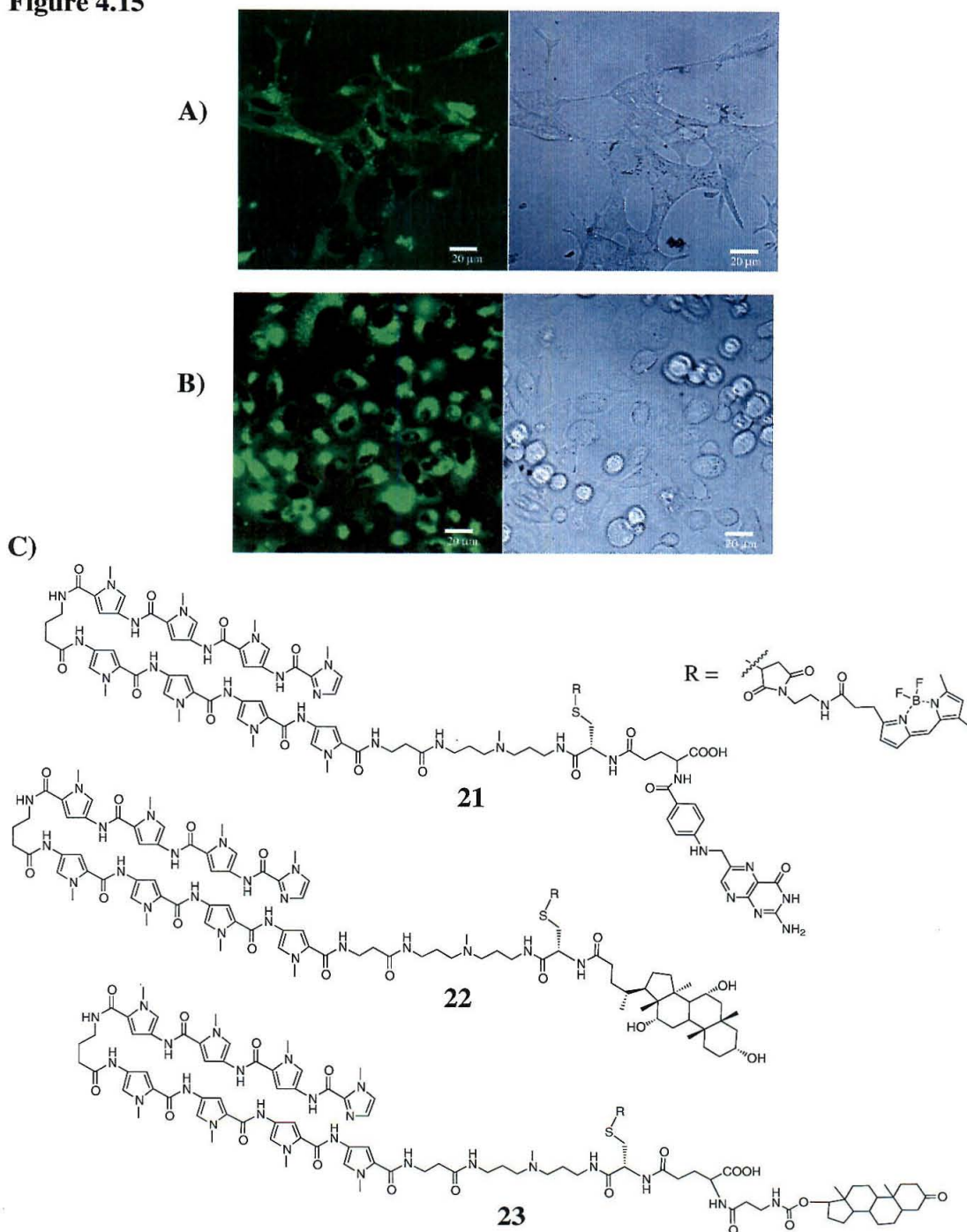
**Table 4.5.** Polyamide-chlorambucil-Bodipy conjugates.

	<b>18</b>	<b>19</b>	<b>20</b>
<b>CEM</b>	nuclear	nuclear	-
<b>Primary T-cells</b>	nuclear	nuclear	-
<b>NB4</b>	nuclear	nuclear	cytoplasm
<b>SKBR-3</b>	cytoplasm	cytoplasm	-
<b>293</b>	cytoplasm	cytoplasm	-
<b>Sf9</b>	cytoplasm	-	-
<b>Kc</b>	cytoplasm	cytoplasm	-
<b>786-O</b>	cytoplasm	cytoplasm	-
<b>Mel</b>	cytoplasm	cytoplasm	-
<b>LnCap</b>	cytoplasm	cytoplasm	-
<b>PC-3</b>	cytoplasm	cytoplasm	-

**Table 4.5.** Localization of polyamide-CHL-Bodipy conjugates **18-19** and non-alkylating analog **20** in live cells as determined by confocal microscopy. The designation “nucleus” indicates observation of fluorescence in the interior of the nucleus. The designation “cytoplasm” indicates cellular, nonnuclear fluorescence. The designation “-“ indicates that the experiment has not yet been performed. Cells were imaged directly following 20-hour incubation with 5  $\mu$ M **18-20** under normal growth conditions for each cell line.

plays a role as well, as greater localization is observed for the eight-ring conjugate **19** than the 2- $\beta$ -2 conjugate **18**. It is interesting to note that in CEM and primary human T-cells, where both **18** and the standard hairpin conjugates are also observed in the nucleus, the alkylator **18** showed significantly greater nuclear localization than **3** (which has the same 2- $\beta$ -2 parent polyamide).

**Small molecule conjugates for receptor-mediated uptake.** Many small molecules such as nutrients and signal transduction messengers are taken up into cells by specific receptors. In many cases, it has been shown that large cargo, including DNA and PNA oligonucleotides,<sup>3,22</sup> and even 31 kD proteins,<sup>4</sup> can be transported into cells via

**Figure 4.15**

**Figure 4.15.** A) Polyamide-DHT conjugate **23** in live LnCap cells. The fluorescent image is on the left, bright field image on the right. Bar: 20  $\mu\text{m}$ . B) DHT conjugate **23** in live PC-3 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20  $\mu\text{m}$ . C) Chemical structures of the small molecule-polyamide-Bodipy conjugates.

**Table 4.6.** Polyamide-Bodipy-small molecule conjugates.

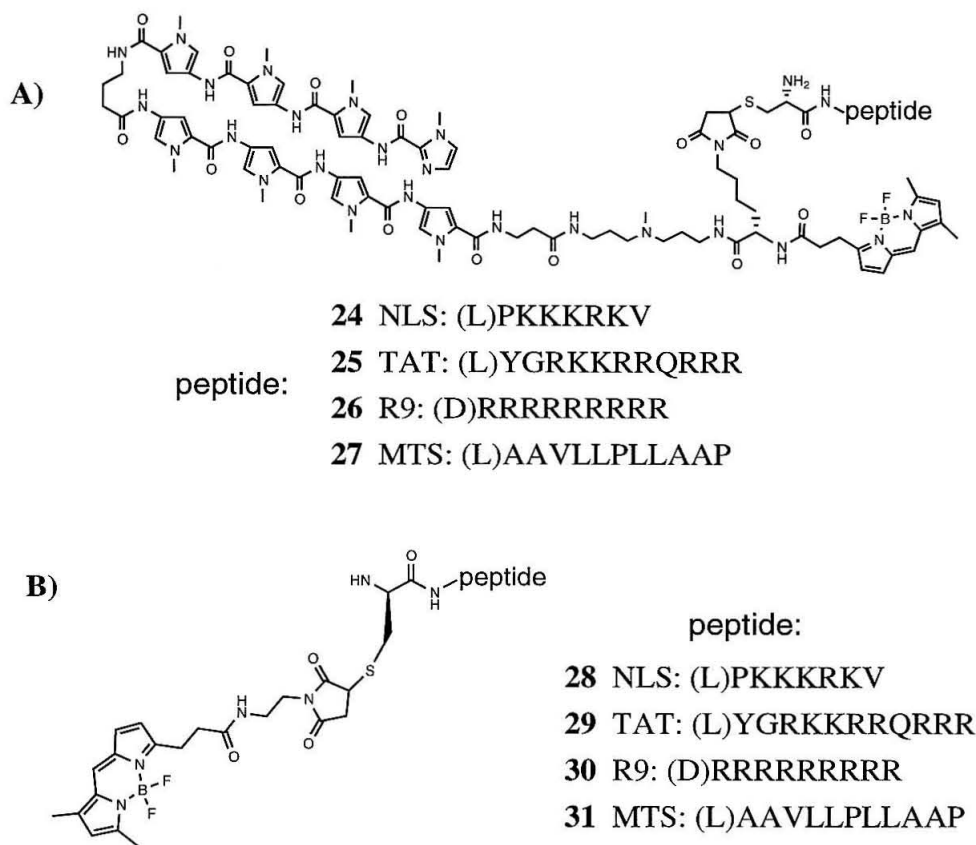
	<b>21</b>	<b>22</b>	<b>23</b>
<b>CEM</b>	-	nuclear	-
<b>NB4</b>	cytoplasm	cytoplasm	cytoplasm
<b>SKBR-3</b>	cytoplasm	cytoplasm	cytoplasm
<b>293</b>	cytoplasm	cytoplasm	cytoplasm
<b>Kc</b>	cytoplasm	cytoplasm	cytoplasm
<b>Mel</b>	cytoplasm	cytoplasm	-
<b>LnCap</b>	cytoplasm	cytoplasm	cytoplasm
<b>PC-3</b>	cytoplasm	cytoplasm	cytoplasm

**Table 4.6.** Localization of polyamide-Bodipy-small molecule conjugates **21-23** in live cells as determined by confocal microscopy. The designation “nucleus” indicates observation of fluorescence in the interior of the nucleus. The designation “cytoplasm” indicates cellular, nonnuclear fluorescence. The designation “-” indicates that the experiment has not yet been performed. Cells were imaged directly following 20-hour incubation with 5  $\mu$ M **21-23** under normal growth conditions for each cell line. For folic acid conjugate **21**, cells were also incubated in folate-free media.

attachment to small molecules in cells containing the cognate receptor. This approach has also been used for small molecule-small molecule conjugates to deliver toxic agents only to specific cells.<sup>23</sup> The folate receptor is often used as a target, since its receptor-mediated endocytosis<sup>24</sup> is well understood and the receptor is overexpressed in a range of cancers.<sup>25</sup> Bodipy-polyamide conjugates with folic acid (**21**), cholic acid (**22**), and dihydrotestosterone (DHT) (**23**) were investigated in a variety of cell lines (Figure 4.15, Table 4.6). The folate experiments were performed with both folate-containing and folate-free media;<sup>4</sup> however, **21** did not successfully reach the nucleus of any cells under either condition. Similarly, DHT conjugate **23** and cholic acid conjugate **22** were ineffective at reaching the nucleus, except for **22** in CEM cells. In the case of prostate cancer cells, which DHT has previously been used to target,<sup>3</sup> there was a marked increase of conjugate in the cytoplasm, for **23** as compared to the standard hairpin-dyes (compare Figure 4.15A, LnCap, to Figure 4.9B). Note the extremely large vesicle-like structures at

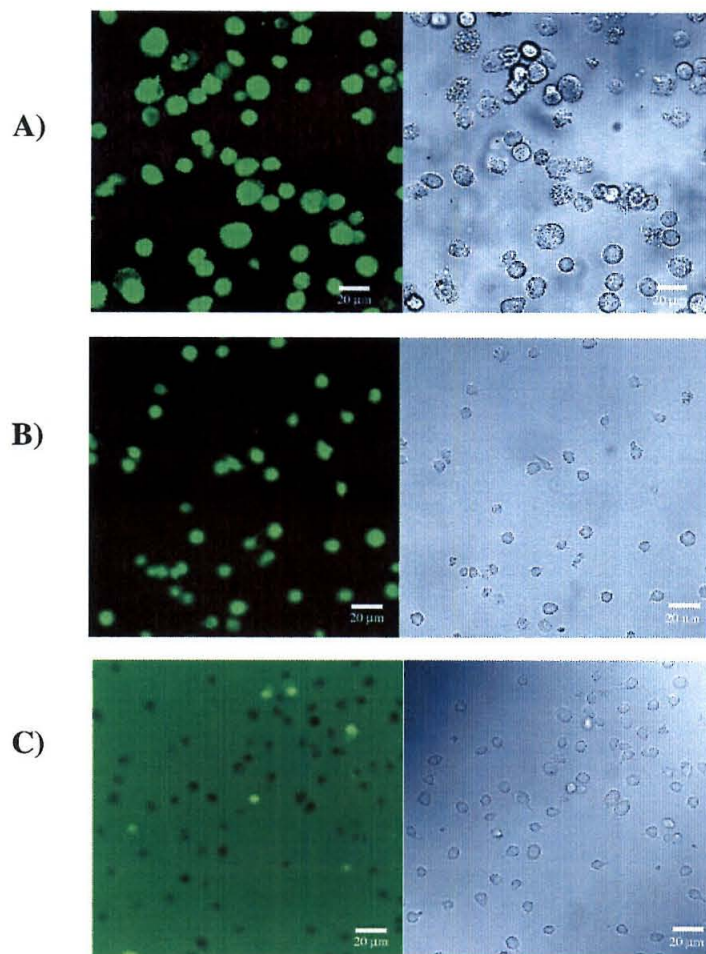
the edges of the PC-3 cells (Figure 4.15B). It appears that DHT is helping the polyamide into the cell, but the conjugate remains trapped in cytoplasm. This is in fact somewhat counter-intuitive, since DHT itself functions in the nucleus.

**Figure 4.16**



**Figure 4.16.** Chemical structures of **(A)** polyamide-carrier peptide-Bodipy conjugates, and **(B)** carrier peptide-Bodipy control compounds.

**Carrier peptide conjugates.** Carrier peptides have been used to transport small molecules, oligonucleotides, and large proteins both into cells and into nuclei,<sup>6</sup> which prompted their investigation in this system. Two of the most widely used carrier peptides are TAT,<sup>7,26-28</sup> a basic domain derived from the HIV-1 Tat protein,<sup>29</sup> and the SV 40 NLS peptide,<sup>30-33</sup> derived from a basic region from the SV40 T antigen protein.<sup>34</sup> While the NLS (nuclear localization signal) has a well-characterized receptor for nuclear import,

**Figure 4.17**

**Figure 4.17.** A) R9 conjugate **26** in live Sf9 cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 µm. B) NLS conjugate **24** in live Kc cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 µm. C) NLS-Bodipy control **28** in live Kc cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 µm.

karyopherin  $\alpha$ ,<sup>34,35</sup> the mechanism of TAT transduction is still unknown despite considerable study. TAT transduction appears not to involve classical receptor-, transporter-, endosome-, or adorative-endocytosis-mediated processes.<sup>6,36</sup> Since it is likely that, at least in some cells, polyamide conjugates are trapped in endosomes, the ability of TAT to bypass endosomes should result in enhanced nuclear localization, especially since TAT itself acts a nuclear localization signal.<sup>37</sup> In addition to NLS (**24**) and TAT (**25**), conjugates of R9 (nine arginines)<sup>8</sup> (**26**) and MTS<sup>38</sup> (**27**) were also

**Table 4.7.** Carrier peptide-polyamide-Bodipy conjugates.

	<b>24</b> NLS	<b>25</b> TAT	<b>26</b> R9	<b>27</b> MTS
<b>CEM</b>	cytoplasm	nuclear	nuclear	cytoplasm
<b>NB4</b>	cytoplasm	nuclear	nuclear	cytoplasm
<b>SKBR-3</b>	cytoplasm	cytoplasm	nuclear*	cytoplasm
<b>MCF-7</b>	cytoplasm	cytoplasm	cytoplasm	cytoplasm
<b>293</b>	cytoplasm	nuclear	nuclear*	nuclear*
<b>Sf9</b>	cytoplasm	cytoplasm	nuclear	cytoplasm
<b>Kc</b>	nuclear	nuclear	nuclear	cytoplasm
<b>Hi5</b>	cytoplasm	cytoplasm	cytoplasm	cytoplasm
<b>S2</b>	nuclear	cytoplasm	cytoplasm	cytoplasm
<b>786-O</b>	cytoplasm	cytoplasm	cytoplasm	cytoplasm

**Table 4.7.** Localization of carrier peptide-polyamide-Bodipy conjugates **24-27** in live cells as determined by confocal microscopy. The designation “nucleus” indicates observation of fluorescence in the interior of the nucleus. The designation “cytoplasm” indicates cellular, non-nuclear fluorescence. The designation “\*” indicates that the compound may be toxic to those cells under these conditions. Cells were imaged directly following 20-hour incubation with 5  $\mu$ M **24-27** under normal growth conditions for each cell line.

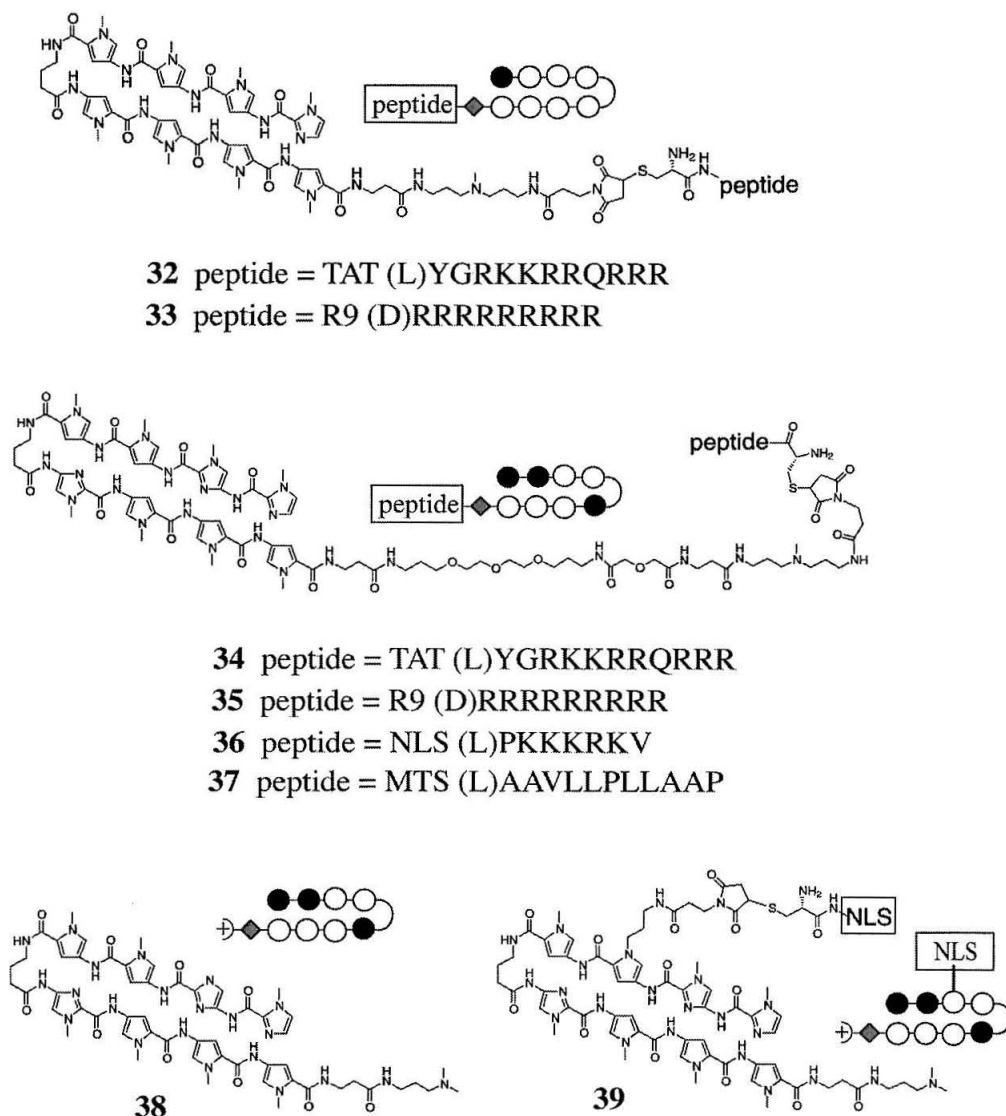
investigated, as well as peptide-Bodipy controls lacking a polyamide (**28-31**) (Figures 4.16 and 4.17, Table 4.7). MTS is a membrane translocation sequence derived from the kaposi FGF signal sequence;<sup>38</sup> it should allow the conjugate to pass through the outer membrane again bypassing endocytosis. R9 was shown by Wender and co-workers, who also investigated a peptoid version, to outperform TAT at uptake into Jurkat cells,<sup>8</sup> but does not necessarily confer nuclear localization. Of note is that of the peptides investigated in this series, only R9 was the D-form, which will limit its potential proteolytic degradation relative to the L-peptides. (Polyamides themselves appear to be resistant to proteases, at least *in vitro*.)<sup>39</sup>

As seen in Table 4.7, the peptide conjugates ranged from nuclear localization in one cell line out of ten examined for the MTS conjugate **27**, to six of out ten for the R9



conjugate **26**. Compound **26** is the best option examined thus far for targeting multiple cell lines (see Figure 4.17A with Sf9 cells), including the human breast cancer cells SKBR-3 and human kidney fibroblasts 293, although it should be noted that, at the concentrations of the assay, 5  $\mu$ M, **26** caused some cell-death in these compounds. The MTS conjugate **27** was also somewhat toxic to 293 cells, the one cell line that it localized to the nucleus in. The TAT (**25**) and NLS (**24**) are intermediate, with **25** reaching the nucleus of CEM, NB4, 293, and the *Drosophila* cell Kc. NLS conjugate **24** bound both *Drosophila* cells (Kc and S2), but only those two cell lines. It is very difficult to speculate about what the actual role of the peptide is in these experiments, as it may be different in different cell lines, and the polyamide conjugates (**24-27**) either equaled or in most cases out-performed the carrier peptide controls (**28-31**) in terms of nuclear localization. One of the biggest surprises in the entire study has been the behavior of the carrier peptide-Bodipy conjugates (**28-31**), which in most cases did not even localize inside the cell (compare 4.17B and 4.17C, compounds **24** and **28** in Kc cells). This was not expected from the literature, particularly for TAT, but on its own, the TAT-Bodipy conjugate went into the nucleus of CEM and was otherwise noncellular. This highlights the difficulty of getting Bodipy into the nucleus. The carrier peptides, while being the best option tested so far, do not impart the extreme generality attributed to them in the literature with regard to cargo and cell type. However, nuclear localization in six out ten cell lines for the R9 conjugate **26**, four out of ten for TAT conjugate **25**, and selectivity for *Drosophila* cells for the NLS conjugate **24**, represented major advances. The next step in preparation for functional *in vivo* assays, DNase I footprinting on polyamide-carrier peptide conjugates (no dye), was eagerly anticipated.

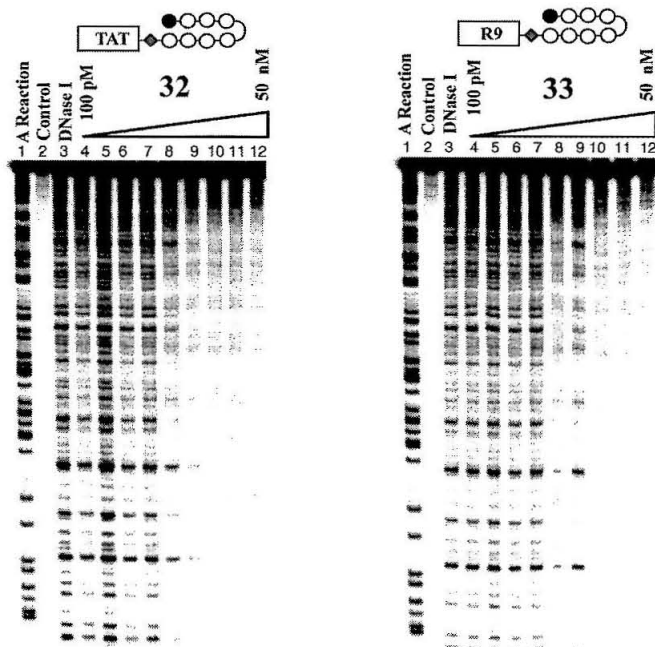
**DNA binding properties of polyamide-carrier peptide conjugates.** The success of the cationic carrier peptides prompted an investigation of the *DNA binding* properties of cationic peptide-polyamide conjugates (no Bodipy). Unfortunately, coating

**Figure 4.18****Figure 4.18.** Structures of compounds examined by DNase I footprinting.

of the DNA (nonspecific binding at all sequences) was consistently observed by quantitative DNase I footprinting titrations.<sup>14</sup> Initially 1-imidazole eight-ring hairpin compounds **32** (TAT) and **33** (R9) (Figure 4.18) based on the 1-imidazole eight ring hairpin conjugates **25** and **26**, respectively, were investigated on the restriction fragment derived from pJT8<sup>15</sup> (Figure 4.19). Next, compounds **34–37** were prepared, with the aim of placing the peptides on a more intrinsically specific polyamide (**38**). A long linker



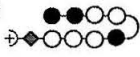
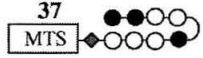


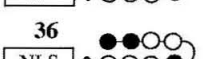
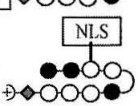
Figure 4.19



**Figure 4.8.** Quantitative DNase I footprint titration experiments on a 3'-<sup>32</sup>P-labeled 229-bp DNA fragment of the plasmid pJT8.<sup>15</sup> All reactions contain 15 kcpm DNA fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. **A)** Lane 1, A reaction; lane 2, intact DNA; lane 3, DNase I standard; lanes 4–12, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, and 50 nM polyamide-TAT peptide conjugate **32**, respectively. **B)** Lane 1, A reaction; lane 2, intact DNA; lane 3 DNase I standard; lanes 4–12, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, and 50 nM polyamide-R9 peptide conjugate **34**, respectively.

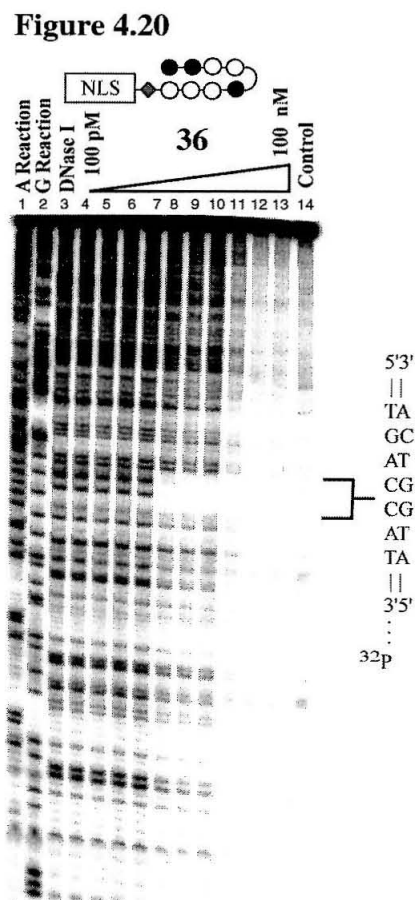
between the polyamide and peptide was also employed in the design of these compounds. DNase I footprinting assays were carried out using the restriction fragment from plasmid pDEH9<sup>40</sup> (Table 4.8, Figure 4.20). The coating phenomenon was still observed with the cationic carrier peptides, which surprisingly had no greater affinity for the match site than the singly charged parent hairpin **38** ( $K_a = 1 \times 10^9 \text{ M}^{-1}$ ). Coating was not observed for conjugate **37** which contains the uncharged, hydrophobic MTS peptide, but this compound bound with 100-fold lower affinity than the parent. In order to confirm that the observed coating was a real property of DNA binding and not the result of an interaction between the cationic peptides and the DNase I enzyme, MPE footprinting experiments<sup>14</sup> were performed which yielded similar results to those shown for DNase I footprinting. The affinity and specificity of the cationic peptides themselves were also

**Table 4.8.** Equilibrium association constants and specificity over coating for polyamide-carrier peptide conjugates.

Polyamide	Match Site $\sim K_a$ ( $M^{-1}$ )	Specificity over coating
38 	$1 \times 10^9$	—
37 	$1 \times 10^7$	—
34 	$1 \times 10^9$	$\sim 5$ -fold
35 	$1 \times 10^9$	$\sim 10$ -fold
36 	$5 \times 10^8$	$\sim 20$ -fold
39 	$1 \times 10^{10}$	$\sim 75$ -fold

**Table 4.8.** Values reported are the mean values obtained from at least three DNase I footprint titration experiments, but are not quantitative, due to the interference of coating. Assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

**Figure 4.20.** Quantitative DNase I footprint titration experiments on a 3'-<sup>32</sup>P-labeled restriction fragment of the plasmid pDEH9.<sup>40</sup> All reactions contain 15 kcpm DNA fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. Lane 1, A reaction; lane 2, G reaction; lane 3, DNase I standard; lanes 4-13, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM and 100 nM polyamide-NLS peptide conjugate **36**, respectively; lane 14, intact DNA.

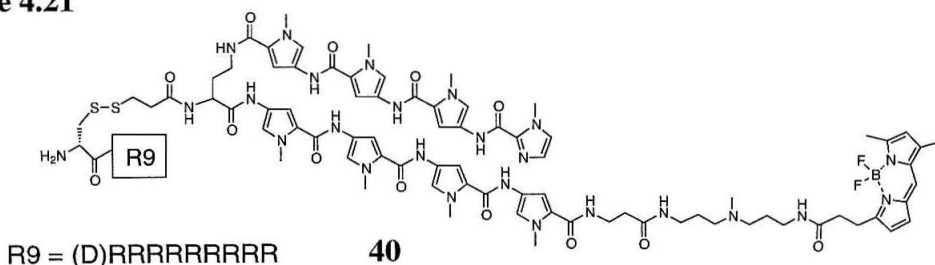


briefly investigated. None were found to exhibit any sequence specificity, and the affinities ranged from  $\sim 10^5 M^{-1}$  for NLS to  $\sim 10^7 M^{-1}$  for R9.<sup>41,42</sup> Consistent with a lower intrinsic affinity for DNA, attachment of NLS to the polyamide afforded the most specific compound (**36**,  $\sim 20$ -fold over coating) of the cationic series. In order to see if the specificity could be further optimized, by moving the attachment point of the NLS from the tail the *N*-methyl position of a Py ring. Indeed both the specificity ( $\sim 75$ -fold over coating) and affinity ( $K_a = 1 \times 10^{10} M^{-1}$ ) of the resulting conjugate (**39**) went up. This compound or other *N*-methyl-linked NLS conjugates are reasonable candidates for *in vivo*

assays in *Drosophila* cells. The specificity is yet not optimal, but may be good enough to see specific effects in cell culture. The analogous compounds with TAT and R9 have not yet been investigated, partially because it is unlikely that the specificity over coating for those compounds would exceed 75-fold.

**Carrier peptide disulfide conjugate.** The loss of DNA binding specificity observed for the most effective carrier peptides is a setback. One strategy for overcoming the problem is a pro-drug approach, where in the optimal case, a cationic carrier peptide is used to bring the polyamide to the nucleus, and is then rapidly cleaved so that the polyamide can go about binding DNA specifically. A common way to attach carrier peptides to their cargo is through a disulfide bond.<sup>6</sup> The carrier peptide brings the cargo into the cell, and in that reductive environment the disulfide bond is cleaved. This differs from the optimal case for polyamides because it is expected that the disulfide bond will be cleaved in the cytoplasm, and then the fate of the polyamide at that point is uncertain. Bodipy conjugate **40**, which contains a disulfide-linked R9, was synthesized to test the utility of this approach (Figure 4.21). If the R9 in conjugate **26** functioned to alter the mechanism of uptake such that the conjugate is “free” in the cytoplasm, the polyamide portion of conjugate **40** should be able to reach the nucleus and the thermodynamic sink of the DNA even with the release of R9 in the cytoplasm. However, if R9 plays an active role in nuclear localization from the cytoplasm, either by interacting

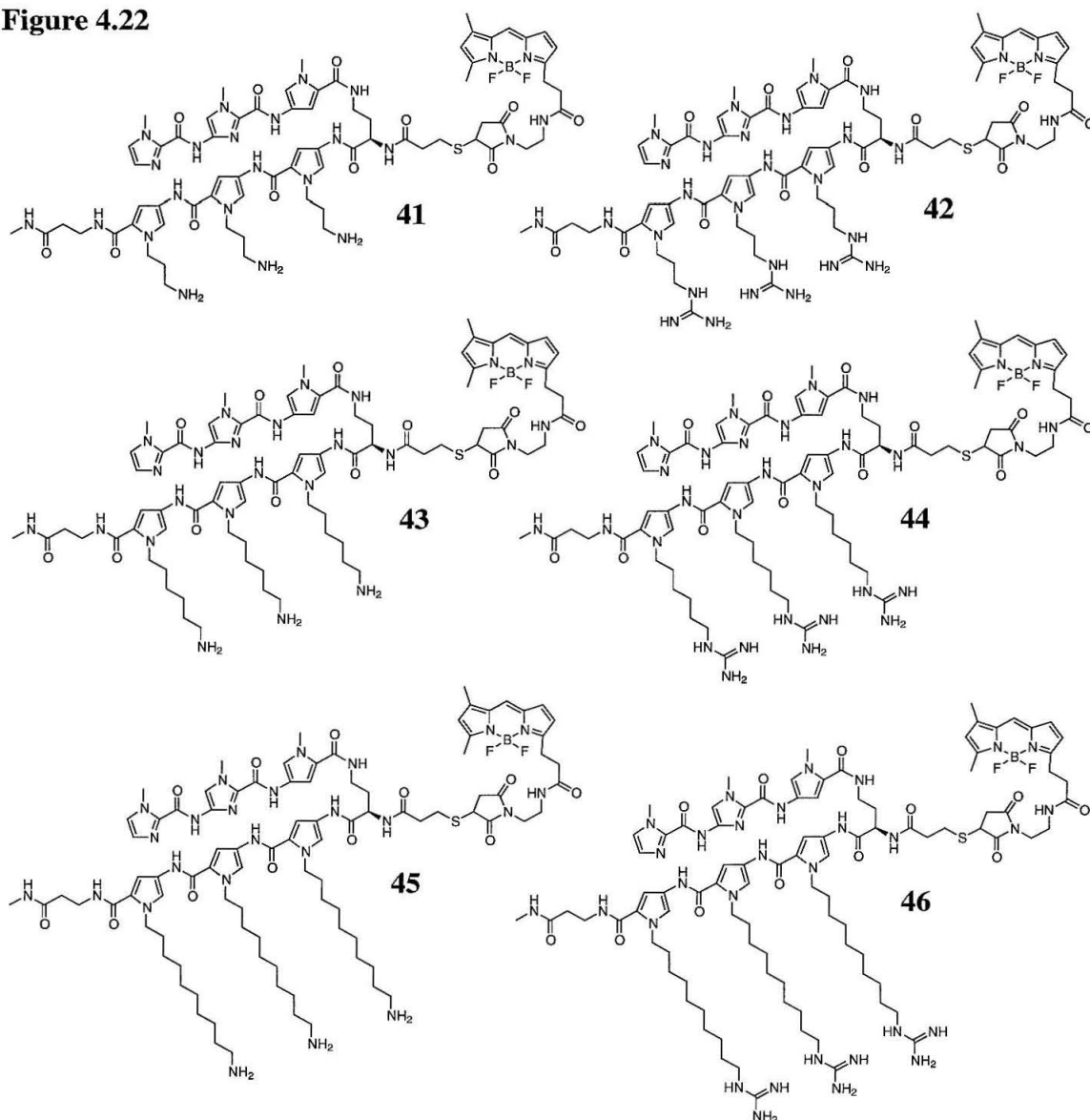
**Figure 4.21**



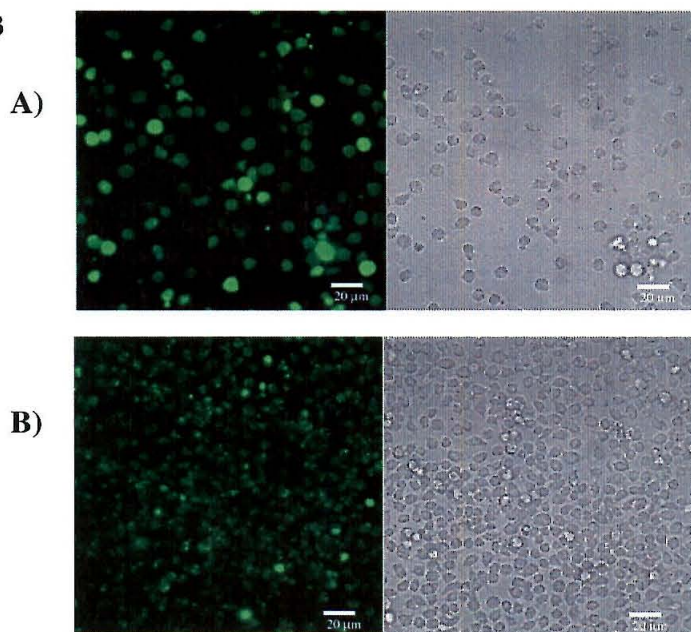
**Figure 4.21.** Chemical structure of the Bodipy-polyamide-disulfide-linked-R9 conjugate **40**.

with a specific nuclear receptor or by changing the properties of the conjugate so that it can escape from vesicles, then the polyamide portion of conjugate **40**, with the Bodipy reporter attached, might be expected to stay in the cytoplasm. Thus, in addition to providing very general solution for trafficking polyamides to the nucleus if successful, conjugate **40** was also expected to be a probe for the mechanism of R9-enhanced nuclear localization. Unfortunately, the disulfide conjugate **40** defied both expectations and was observed only very weakly even in the cytoplasm of all the cell lines tested (primarily noncellular). Despite the failure of this particular compound, a pro-drug type of approach may still be successful, although other common options such as ester linkages to be cleaved by esterases, share the original concern of the disulfide, in that the ester linkage would be cleaved in the cytoplasm.

**Cationic polyamide-Bodipy conjugates.** A common feature of the successful carrier peptides was their cationic charge, and simplistically, the more charges the better. This is indeed what Wender et al. observed quantitatively for uptake into Jurkat cells, varying from five to nine guanidines.<sup>8</sup> For peptides the D-arginine oligomer was preferable to the L-arginine oligomer, and in a peptoid series, the longer alkyl chain length between the peptoid backbone and the guanidine unit, the better. Their results only concerned cellular uptake in a single cell line and not nuclear localization. However, since in our case positively charged peptides were beneficial for nuclear localization, but were a problem for specificity, replacement of the entire carrier peptide by charged groups on the polyamide backbone was investigated. This work is in analogy to the Wender peptoids,<sup>8</sup> and builds on previous work in the Dervan laboratory investigating “positive patch” polyamides.<sup>43,44</sup> Quantitative DNase I footprinting experiments have revealed that six-ring hairpins with either a primary amine or a guanidine and a linker of three or six carbon atoms connected to the *N*-methyl carbon of three of the pyrrole units show good specificity.<sup>45</sup> Corresponding compounds with ten

**Figure 4.22****Figure 4.22.** Chemical structures of Bodipy-cationic polyamide conjugates.

carbon linkers were also investigated, but showed poorer specificity. Bodipy conjugates (41-46) of each member of the series were prepared and examined by confocal microscopy (Figure 4.22 and 4.23, Table 4.9). Given that the six-rings have only three charges, the observation that some members of the series localize to the nucleus of three cell lines (NB4, CEM, and Kc) was highly promising. While the entire series localized to the nucleus of CEM cells, the conjugates that localized to the nucleus of NB4 cells did

**Figure 4.23**

**Figure 4.23.** A) Cationic polyamide conjugate **41** in live CEM cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 µm. B) Cationic polyamide conjugate **41** in live Kc cells. The fluorescent image is on the left, bright field image on the right. Bar: 20 µm.

**Table 4.9.** Cationic polyamide-Bodipy conjugates.

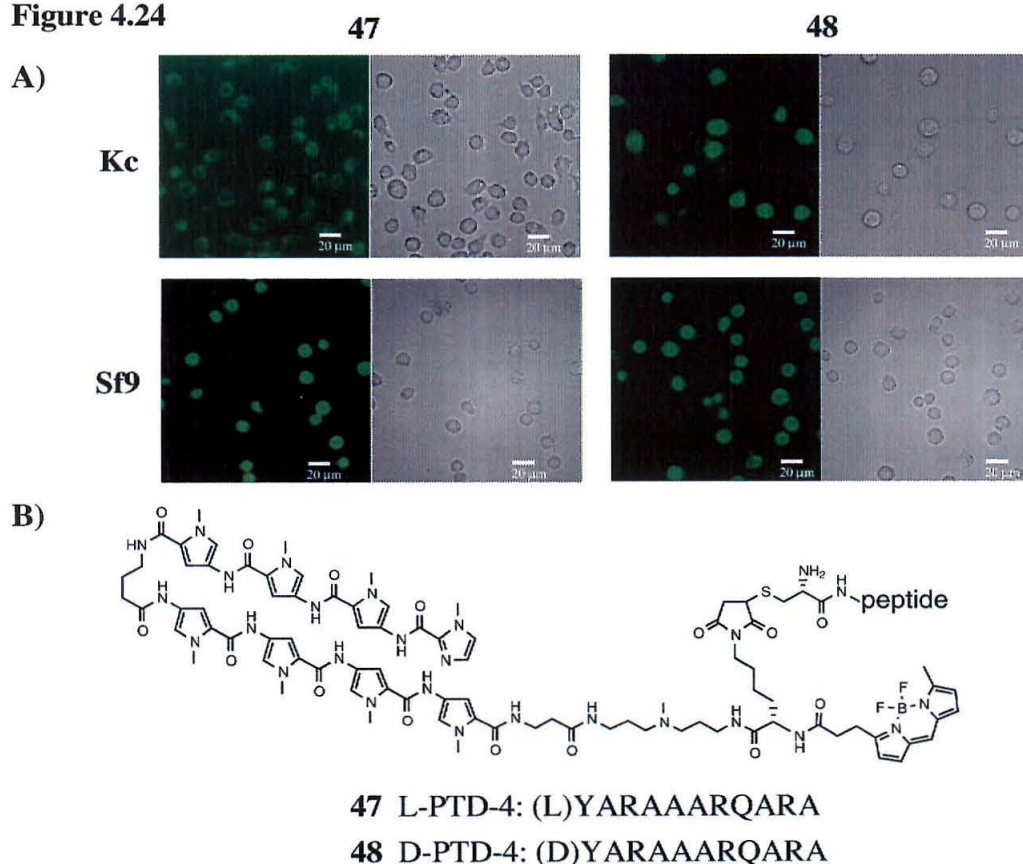
	<b>41</b>	<b>42</b>	<b>43</b>	<b>44</b>	<b>45</b>	<b>46</b>
<b>CEM</b>	nuclear	nuclear	nuclear	nuclear	nuclear	nuclear
<b>NB4</b>	cytoplasm	#	cytoplasm	nuclear	nuclear	cytoplasm
<b>SKBR-3</b>	-	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm
<b>Mel</b>	-	cytoplasm	cytoplasm	cytoplasm	cytoplasm	-
<b>293</b>	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm
<b>Sf9</b>	cytoplasm	cytoplasm	nuclear	cytoplasm	cytoplasm	#
<b>Kc</b>	nuclear	nuclear	nuclear	cytoplasm	cytoplasm	cytoplasm
<b>LnCap</b>	-	cytoplasm	cytoplasm	cytoplasm	cytoplasm	-
<b>PC3</b>	-	cytoplasm	cytoplasm	cytoplasm	cytoplasm	-

**Table 4.9.** Localization of cationic polyamide-Bodipy conjugates **41-46** in live cells as determined by confocal microscopy. The designation “nucleus” indicates observation of fluorescence in the interior of the nucleus. The designation “cytoplasm” indicates cellular, nonnuclear fluorescence. The designation “#” indicates inconsistent results, and the designation “-” indicates the experiment has not been performed yet. Cells were imaged directly following 20-hour incubation with 5 µM **41-46** under normal growth conditions for each cell line.



not localize to the nucleus of Kc cells, and vice versa with the possible exception of **42** (for which the results are currently inconsistent). The good specificity of the compounds with three- and six-carbon linkers suggest that these cationic hairpin polyamides would be useful in NB4, Kc, or CEM cells for *in vivo* experiments. Currently, efforts are underway to examine cationic eight-ring hairpin-Bodipy conjugates with three, four, and five charges with the goal of localization to the nucleus of human cell lines such as 293 and SKBR-3, which were amenable to nuclear localization of the TAT and R9 conjugates.

**Figure 4.24**



**Figure 4.15.** A) PTD-4 conjugates **47** and **48** in live Kc and Sf9 cells. Each fluorescent image is on the left, bright field image on the right. Bar: 20  $\mu$ m. B) Chemical structures of the L- and D-PTD-4 polyamide-Bodipy conjugates **47** and **48**, respectively.



**Additional carrier peptide conjugates.** Concurrent with the efforts to build more positive charges into the polyamide backbone, carrier peptides with fewer positive charges were also investigated. Dowdy and co-workers have recently reported an alanine-rich derivative of TAT called PTD-4 (protein transduction domain four) which only has three of the original eight charges.<sup>46</sup> It was shown to be effective at transducing into cells both in culture (Jurkat) and in mice. Polyamide-Bodipy conjugates with both the L (47) and D (48) peptides were prepared and investigated by confocal microscopy. Figure 4.24 shows that there is a functional difference between the L and D forms in Kc cells, only the latter (48) is observed in the nucleus at the 20-hour imaging point. In Sf9 cells, both were observed in the nucleus. Unfortunately, these were the only cell lines where either 47 or 48 were found in the nucleus (out of over ten cell lines tested). This is particularly unfortunate because very recent DNase I footprinting experiments have shown that a PTD-4 conjugate has good specificity over coating. DNase I footprinting experiments have also recently been performed for truncated versions of R9. Uptake of R5 and R7 peptides has been demonstrated in Jurkat cells,<sup>8</sup> and R7 has been used to transport a cell-impermeable peptide into cells and intact heart.<sup>47</sup> While the R7 conjugate had displayed coating similar to R9 conjugates 33 and 35, the R5 conjugate showed better specificity, on the order of NLS. The very promising R5 conjugate is currently under investigation by confocal microscopy.

## Conclusions

At the outset of these studies, nuclear localization of standard polyamide-dye conjugates had been observed in two related cell lines. Two others were uncovered recently, and all four are related to the human immune system. This clearly represents an important lead that will be followed up in the future. Currently, the efforts described in this chapter have yielded six additional cell lines ranging from *Drosophila* to human breast cancer, amenable to nuclear localization of certain polyamide derivatives. Among

the modifications tested, small molecule conjugates for receptor-mediated uptake were the least successful, and cationic carrier peptides, particularly TAT and R9, were the most successful. The majority of the recent efforts have been focused on exploring the complex relationship between positive charges (either appended as carrier peptides or attached directly to the polyamide backbone), nuclear localization, and DNA binding specificity. However, simply moving to a smaller conjugate with the one-to-one binders yielded the same number of positives as TAT. This is also clearly a direction that will be followed up in the future, and a molecular weight study has been initiated. Yet, it is important to note that it is mainly the same cell lines that are accepting very different conjugates. Multiple solutions have now been found for CEM, NB4, Sf9, and Kc. Fundamentally different approaches may be necessary to gain access to the nucleus of the more recalcitrant cells. Confocal microscopy will inform functional assays and drive this ongoing discovery process.

## Experimental

**Polyamide Conjugates.** Compounds **2,19-37,39-40**, and **47-48** were synthesized by Bobby Arora. Compounds **41-46** were synthesized by Ben Edelson and Amanda Cashin. Compounds **13-14** were synthesized by Michael Marques and Adam Urbach. Compound **18** was synthesized by Nick Wurtz and compound **12** was synthesized by Philipp Weyermann. Compound **4** was originally synthesized by Roland Burli and subsequently by the author. **1**: [M+H] 1538.9, 1538.7 calculated for [M+H]. **3**: [M+H] 1683.1, 1383.8 calculated for [M+H]. **4**: [M+H] 1783.8, 1783.7 calculated for [M+H]. **5**: [M+H] 1681.0, 1680.8 calculated for [M+H]. **6**: [M+H] 1731.1, 1731.8 calculated for [M+H]. **7**: [M+H] 1539.0, 1539.7 calculated for [M+H]. **8**: [M+H] 777.5, 777.4 calculated for [M+H]. **9**: [M+H] 899.6, 899.5 calculated for [M+H]. **10**: [M+H] 1921.0, 1920.9 calculated for [M+H]. **11**: [M+Na] 1703.1, 1702.6 calculated for [M+Na].

**Quantitative DNase I Footprinting.** As previously reported,<sup>14</sup> for DNA fragments derived from the HER2 promoter,<sup>9</sup> plasmid pJT8,<sup>15</sup> and plasmid pDEH9.<sup>40</sup>

**Confocal Microscopy under live cell conditions.** As previously reported in Chapter 4A.

## Acknowledgements

We thank the NIH for research support and the Ralph M. Parsons Foundation for a predoctoral fellowship to J.M.B. We are grateful to everyone who contributed to this project, most notably Dr. Bobby Arora, Dr. Stephanie Leslie, Nick Nickols, Ben Edelson, Inder Nangiana, Dr. Leonard Prins, Amanda Cashin, and those already mentioned for synthesis.

## References and Notes

- (1) Compound numbering follows from Chapter 4A.
- (2) Wurtz, N. R.; Dervan, P. B. *Chem. Biol.* **2000**, *7*, 153-161.
- (3) Boffa, L. C.; Scarfi, S.; Mariani, M. R.; Damonte, G.; Allfrey, V. G.; Benatti, U.; Morris, P. L. *Cancer Res.* **2000**, *60*, 2258-2262.
- (4) Leamon, C. P.; Low, P. S. *J. Biol. Chem.* **1992**, *267*, 24966-24971.
- (5) Zhang, X.; Simmons, C. G.; Corey, D. R. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1269-1272.
- (6) Fischer, P. M.; Krausz, E.; Lane, D. P. *Bioconjugate Chem.* **2001**, *12*, 825-841.
- (7) Schwarze, S. R.; Ho, A.; Vocero-Akbani, A.; Dowdy, S. F. *Science* **1999**, *285*, 1569-1572.
- (8) Wender, P. A.; Mitchell, D. J.; Pattabiraman, K.; Pelkey, E. T.; Steinman, L.; Rothbard, J. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13003-13008.

- (9) Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A. *J. Biol. Chem.* **2000**, *275*, 24246-24254.
- (10) Leslie, S. J. Ph.D. Dissertation, Rosewell Park Cancer Institute: Buffalo, NY, 2001.
- (11) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895.
- (12) Rucker, V.; Foister, S.; Dervan, P.B. Unpublished results.
- (13) Rucker, V.; Foister, S.; Chevillet, J.R.; Dervan, P.B. Unpublished results.
- (14) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.
- (15) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561.
- (16) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121-1129.
- (17) Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem.-Eur. J.* **1999**, *5*, 975-983.
- (18) Urbach, A. R.; Dervan, P. B. *Proc Natl Acad Sci U S A* **2001**, *98*, 4343-4348.
- (19) Janssen, S.; Durussel, T.; Laemmli, U. K. *Mol. Cell* **2000**, *6*, 999-1011.
- (20) Janssen, S.; Cuvier, O.; Muller, M.; Laemmli, U. K. *Mol. Cell* **2000**, *6*, 1013-1024.
- (21) Wurtz, N. R. Ph.D. Dissertation, California Institute of Technology: Pasadena, CA, 2002.
- (22) Zhang, X.; Simmons, C. G.; Corey, D. R. *Bioorg. Med. Chem.* **2001**, *11*, 1269-1272.
- (23) Steinberg, G.; Borch, R. F. *J. Med. Chem.* **2001**, *44*, 69-73.
- (24) Leamon, C. P.; Low, P. S. *Proc Natl Acad Sci USA* **1991**, *88*, 5572-5576.
- (25) Atkinson, S. F.; Bettinger, T.; Seymour, L. W.; Behr, J.-P.; Ward, C. M. *J. Biol. Chem.* **2001**, *276*, 27930-27935.

- (26) Wunderbaldinger, P.; Josephson, L.; Weissleder, R. *Bioconjugate Chem.* **2002**, *13*, 264-268.
- (27) Lee, H. J.; Pardridge, W. M. *Bioconjugate Chem.* **2001**, *12*, 995-999.
- (28) Polyakov, V.; Sharma, V.; Dahlheimer, J. L.; Pica, C. M.; Luker, G. D.; Piwnicka-Worms, D. *Bioconjugate Chem.* **2000**, *11*, 762-771.
- (29) Mann, D. A.; Frankel, A. D. *EMBO J.* **1991**, *10*, 1733-1739.
- (30) Benimetskaya, L.; Guzzo-Pernell, N.; Lui, S.-T.; Lai, J. C. H.; Miller, P.; Stein, C. A. *Bioconjugate Chem.* **2002**, *13*, 177-187.
- (31) Cutrona, G.; Carpaneto, E. M.; Ulivi, M.; Roncella, S.; Landt, O.; Ferrani, M.; Boffa, L. C. *Nature Biotech.* **2000**, *18*, 300-303.
- (32) Brandem, L. J.; Mohamed, A. J.; Smith, C. I. V. *Nature Biotech.* **1999**, *17*, 784-787.
- (33) Zanta, M. A.; Belguise-Valladier, P.; Behr, J.-P. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 91-96.
- (34) Chook, Y. M.; Blobel, G. *Curr. Opin. Struct. Biol.* **2001**, *11*, 703-715.
- (35) Conti, E.; Uy, M.; Leighton, L.; Blobel, G.; Kuriyan, J. *Cell* **1998**, *94*, 193-204.
- (36) Schwarze, S. R.; Dowdy, S. F. *Trends Pharm. Sci.* **2000**, *21*, 45-48.
- (37) Efthymiadis, A.; Briggs, L.; Jans, D. *J. Biol. Chem.* **1998**, *273*, 1623-1628.
- (38) Rojas, M.; Donahue, J. P.; Tan, Z.; Lin, Y.-Z. *Nature Biotech.* **1998**, *16*, 370-375.
- (39) Arora, P.S.; Dervan, P.B. Unpublished results.
- (40) White, S. Ph.D. Dissertation, California Institute of Technology: Pasadena, CA, 1999.
- (41) Polyarginine and shorter oligomers such as R9 have been investigated as transfection reagents due to their affinity for DNA. See reference 42.
- (42) Futaki, S.; Ohashi, W.; Suzuki, T.; Niwi, M.; Tanaka, S.; Ueda, K.; Harashima, H.; Sugiura, Y. *Bioconjugate Chem.* **2001**, *12*, 1005-1011.
- (43) Bremer, R. E.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1998**, *5*, 119-133.

- (44) Bremer, R. E.; Wurtz, N. R.; Szewczyk, J. W.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2093-2103.
- (45) Cashin, A.; Edelson, B. S.; Dervan, P.B. Unpublished results.
- (46) Ho, A.; Schwarze, S. R.; Mermelstein, S. J.; waksman, G.; Dowdy, S. F. *Cancer Res.* **2001**, *61*, 474-477.
- (47) Chen, L.; Wright, L. R.; Chen, C.-H.; Oliver, S. F.; Wender, P. A.; Mochly-Rosen, D. *Chem. Biol.* **2001**, *8*, 1123-1129.

## CHAPTER 5

### **Synthesis and Investigation of DNA Binding Polyamides with New C-Terminal Tails**

*The text of this chapter was taken in part from a manuscript coauthored with Doan Nguyen, Nick Wurtz, and Professor Peter Dervan (Caltech).*

(Belitsky, J.M.; Nguyen, D.H.; Wurtz, N.R. Dervan, P.B. “Solid Phase Synthesis of DNA Binding Polyamides on Oxime Resin” *Boorg. Med. Chem.*, **in press**)



## Abstract

Control of the energetics and specificity of DNA binding polyamides is necessary for inhibition of protein-DNA complex formation and gene regulation studies. Typically, solid phase methods using Boc monomers for synthesis have depended on Boc- $\beta$ -Ala-PAM resin which affords a  $\beta$ -alanine-Dp tail at the C-terminus, after cleavage with *N,N*-dimethylaminopropylamine (Dp). In Chapter 5A, the energetic consequences of this tail for DNA minor-groove binding are addressed through the synthesis of polyamides with incrementally shortened C-terminal tails. These polyamides were synthesized using an alternative solid-phase method employing the Kaiser oxime resin. Polyamides without Dp and having methyl amide tails rather than  $\beta$ -alanine show similar affinity relative to the standard  $\beta$ -Dp tail. The truncated tails diminish the A,T base pair energetic preference of the  $\beta$ -Dp tail which will allow a greater variety of DNA sequences to be targeted by hairpin polyamides.

Chapter 5B describes an alternate solid-phase route to one of the truncated tails, the primary amide. The aminooxy linker employed in this route was derived from a novel compound generated for a project aimed at a G-selective  $\beta$ -alanine analog. During the development of the aminooxy route to C-terminal pyrrole primary amides, an unexpected reaction with cyclic secondary amides was uncovered, which provides polyamides with tertiary amide tails in moderate yield. The C-terminal cyclic tertiary-amide polyamides are effective DNA binding agents worthy of further development, and have potential as inhibitors of TBP and ESX binding to the HER2 promoter.

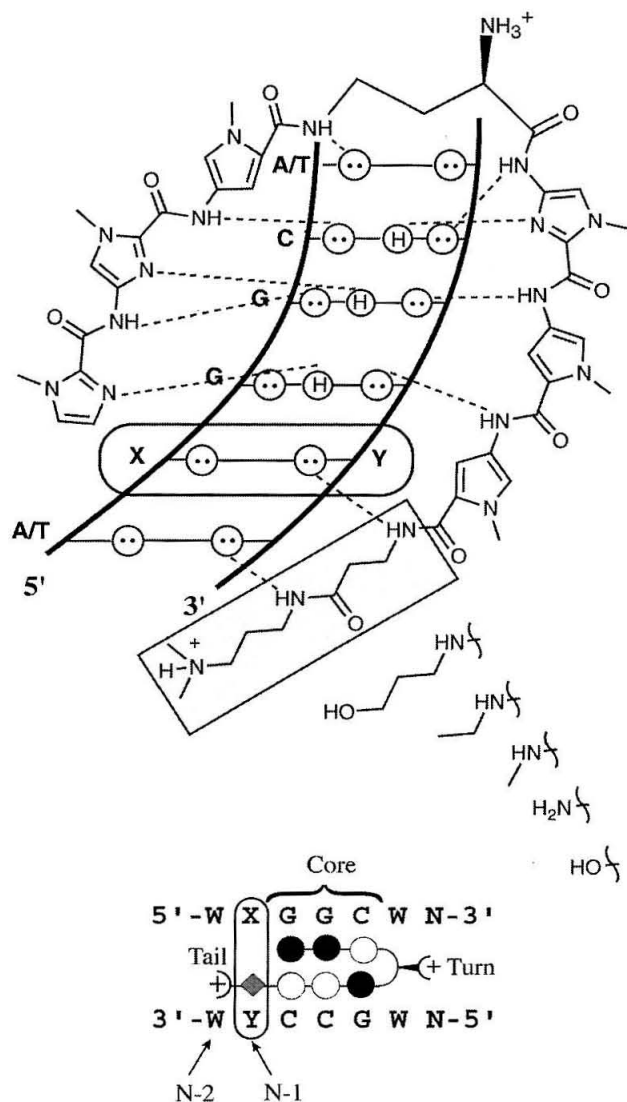
## Chapter 5A:

### Solid-Phase Synthesis of DNA Binding Polyamides on Oxime Resin

#### Introduction

Hairpin polyamides containing *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methyl-3-hydroxypyrrole (Hp) residues bind specific predetermined sequences in the minor groove of DNA with affinities and specificities comparable to naturally occurring DNA-binding proteins.<sup>1</sup> DNA recognition depends on the side-by-side amino acid pairings in the minor groove that stack the aromatic rings against each other and the walls of the groove allowing backbone amide hydrogens and the substituents at the 3-position of the Py, Im, and Hp residues to make specific contacts with the edges of the DNA base pairs. An Im/Py and Py/Im specifies for G•C and C•G, respectively. A Py/Py pair binds A•T and T•A and a Hp/Py pair discriminates T•A over A•T base pairs. These modular ring pairs may be considered the “core” of the polyamide recognition motif. For hairpins an alkyl amino acid, either  $\gamma$ -aminobutyric acid (GABA,  $\gamma$ ), or the chiral, amine-functionalized derivative (*R*)-2,4,-diaminobutyric acid (DABA, (*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ ), serves as the covalent linker region (referred to as the “turn”) between the N-terminal and C-terminal strands and is specific for A•T and T•A base pairs. The “tail” or C-terminal portion of a typical hairpin polyamide is related to the method of synthesis and comprises  $\beta$ -alanine ( $\beta$ ) and *N,N*-dimethylaminopropylamine (Dp), which specify for A,T base pairs at the N-1 and N-2 positions (Figure 5.1).<sup>1</sup>

The  $\beta$ -Dp tail originated with the C $\rightarrow$ N synthesis of polyamides by solid-phase methods using Boc monomers on Boc- $\beta$ -alanine-PAM-resin.<sup>2,3</sup> An alkyl primary amine such as Dp affords efficient cleavage from the solid support by aminolysis of the  $\beta$ -alanine ester linkage to the resin. The sequence requirements of the  $\beta$ -Dp tail for A,T versus G,C at the N-1 and N-2 positions<sup>4</sup> presumably results from a steric clash with the exocyclic NH<sub>2</sub> of a G,C base pair on the floor of the minor groove of DNA (Figure 5.1).

**Figure 5.1**

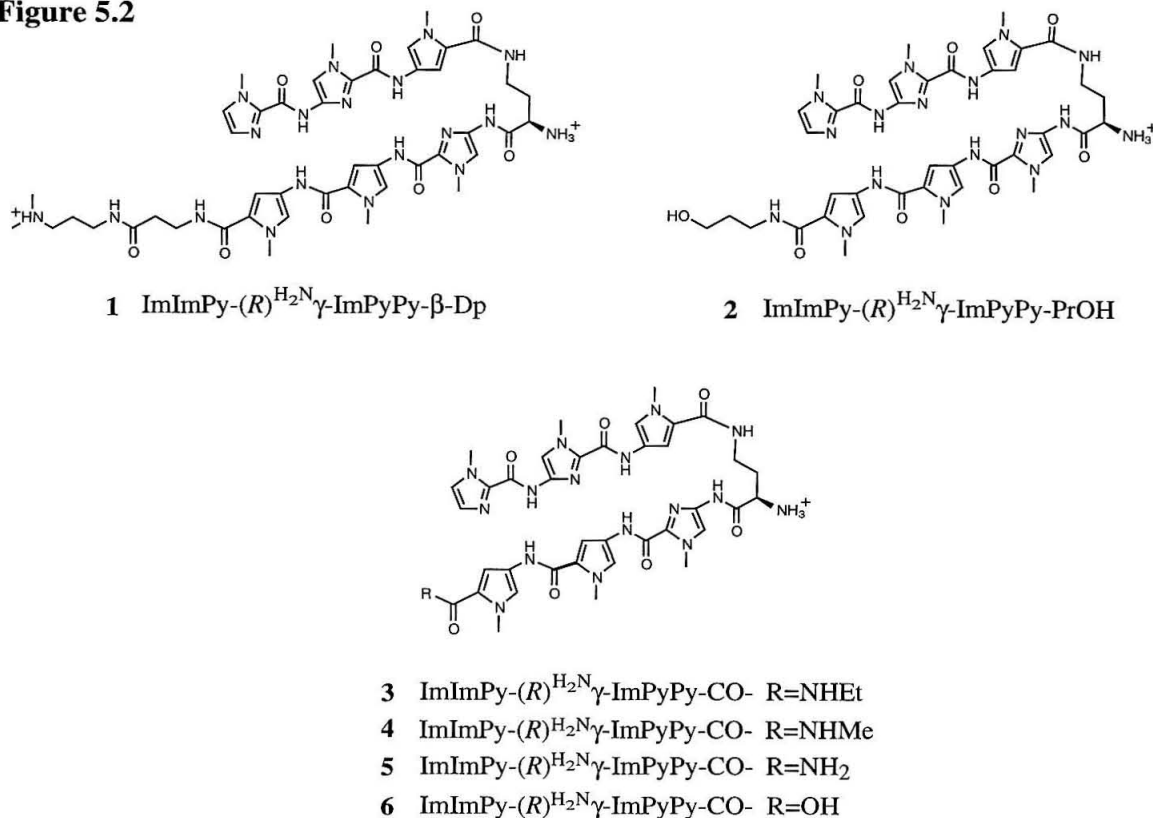
**Figure 5.1.** Binding model for polyamide **1** with DNA. (Top) Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanine. Putative hydrogen bonds are illustrated by dotted lines. The series of truncated tails to be examined are denoted. (Bottom) A ball-and-stick representation of polyamide **1** with DNA. Filled circles denote imidazole while open circles represent pyrrole. The diamond represents  $\beta$ -alanine. (*R*)-diaminobutyric acid ( $(R)^{H_2N}\gamma$ ) and dimethylaminopropylamine (Dp) are depicted as a curved line and a plus sign, respectively. W signifies A or T, N represents any nucleotide. The N-2, N-1, core, turn positions for the polyamide/DNA complex are defined in the text.

Replacement of the  $\beta$ -Dp tail with a shorter propanol amide, generated by the reductive cleavage of the  $\beta$ -alanine-PAM ester from the solid support,<sup>5</sup> removes the preference for an A,T base pair at the N-2 position while maintaining the binding affinity for these sites,

despite the loss of an energetically favorable positive charge.<sup>4</sup> This suggests that the Dp tail interaction with the minor groove may be sterically unfavorable. A deeper understanding of the effect of incremental changes to the tail on the energetics of DNA binding would provide insight on the molecular determinant for A,T versus G,C specificity at the N-1 position.

In order to incrementally reprogram the polyamide C-terminus in the minor groove of DNA, new solid-phase methodology would be required since the propanol amide tail is the shortest C-terminus that can be generated from cleavage of Boc- $\beta$ -Ala-PAM resin. Since the reductive cleavage to generate the propanol amide tail results in poor overall recovery, we generally do not use it. Therefore, for high-affinity recognition, hairpin polyamides are synthesized with  $\beta$ -Dp tails which require DNA binding sites to be composed of the sequence 5'-WW(N)<sub>x</sub>W-3' (W = A,T; N = A,T,G,C; and x = the number of ring pairs in the core of the polyamide). While a large number of biologically relevant sequences have been targeted by the current generation of hairpin polyamides,<sup>6</sup> the necessity for A,T base pairs flanking the polyamide core is a limitation. Promoters are known to have high G,C content associated with CpG islands,<sup>7</sup> and many transcription factors bind G,C rich sequences.<sup>8</sup> In order to effectively inhibit these transcription factors, polyamides that can readily target all base pairs at the tail positions with high affinity are desirable.

We describe here the solid-phase synthesis of polyamides with a pyrrole unit directly linked to the Kaiser oxime resin.<sup>9,10,11</sup> Six-ring polyamide hairpins with the common core sequence ImImPy-(R)<sup>H<sub>2</sub>N</sup> $\gamma$ -ImPyPy-X (**3-6**) were synthesized where X represents four different C-terminal tails, ethyl amide, methyl amide, primary amide, and carboxylic acid (Figure 5.2). The DNA-binding affinity and specificity at the N-1 position was compared to the parent polyamides having either a  $\beta$ -Dp (**1**) or propanol amide tail (**2**).

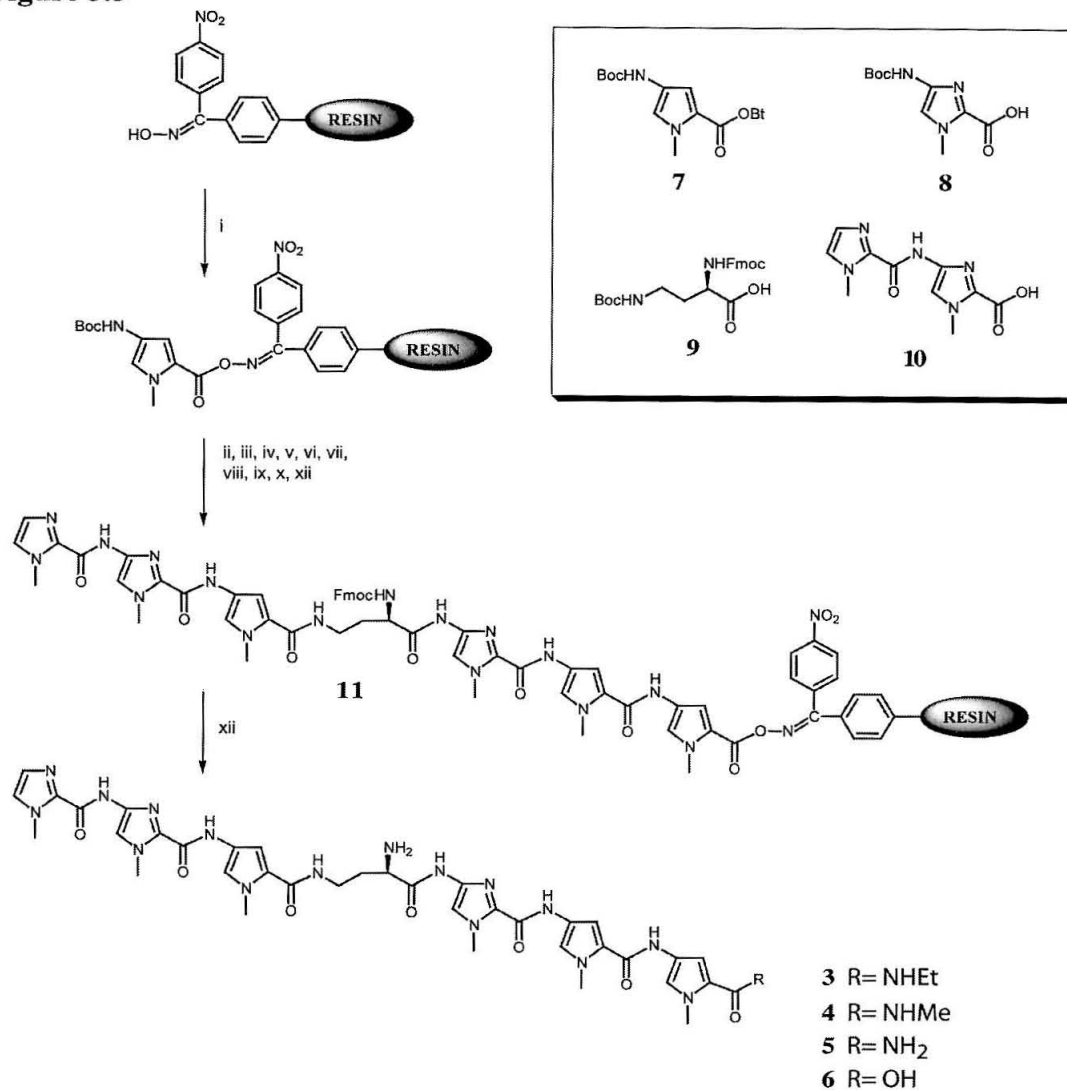
**Figure 5.2**

**Figure 5.2.** Structures of polyamides ImImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-β-Dp (**1**), ImImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-PrOH (**2**), ImImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-CONHEt (**3**), ImImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-CONHMe (**4**), ImImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-CONH<sub>2</sub> (**5**) and ImImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-CO<sub>2</sub>H (**6**).

## Results

**Polyamide Synthesis.** Polyamides ImImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-β-Dp (**1**) and ImImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-PrOH (**2**) were synthesized by solid-phase methods on the usual Boc-β-alanine-PAM-resin. Four new polyamides **3-6** were synthesized by solid-phase methods on the Kaiser oxime resin with standard Boc-pyrrole and Boc-imidazole monomers (Figure 5.3). Developed by Kaiser and DeGrado,<sup>9,10</sup> the oxime resin is a versatile polystyrene solid support that is amenable to Boc chemistry and allows for the synthesis of a variety of carboxylic acid derivatives by nucleophilic cleavage from the resin. Peptides can be cleaved from oxime resin to yield primary amide,<sup>10</sup> alkyl amide,<sup>12</sup> and carboxylic acid<sup>13,14</sup> C-termini. The oxime linker is reported to be somewhat acid

Figure 5.3



**Figure 5.3.** Synthesis of polyamide **3-6** from oxime resin using Boc-protected monomers: (i) **7**, DIEA, NMP; (ii) 20% TFA/DCM; (iii) **7**, DIEA, NMP; (v) 20% TFA/DCM; (v) **8**, HBTU, DIEA, NMP; (vi) 50% TFA/DCM; (vii) **9**, HBTU, DIEA, NMP, 37° C; (viii) 20% TFA/DCM; (ix) **7**, DIEA, NMP; (x) 20% TFA/DCM; (xi) **10**, HBTU, DIEA, NMP; (xii) a)  $\text{R} = \text{NHEt}$ , 2.0 M  $\text{CH}_3\text{CH}_2\text{NH}_2/\text{THF}$ , DCM, 37° C; b)  $\text{R} = \text{NHMe}$ , 2.0 M  $\text{CH}_3\text{NH}_2/\text{THF}$ , DCM, 37° C; c)  $\text{R} = \text{NH}_2$ ,  $\text{NH}_3/\text{THF}$ , DCM, DBU, 37° C; or d)  $\text{R} = \text{OH}$ ,  $\text{H}_2\text{O}$ , DMF, DBU, 37° C.

labile, and solutions of less than 25% TFA are recommended for the deprotection of Boc groups. Consequently, a 20% TFA/ $\text{CH}_2\text{Cl}_2$  solution was used for 30 minutes to remove the Boc groups from pyrrole and aliphatic amines. However, it was found that Boc-imidazole residues could not be fully deprotected under these conditions. Longer

deprotection times resulted in increasing degradation of the resin bound polyamide. Boc groups could be efficiently removed from imidazoles in 30 minutes with a 50% TFA/CH<sub>2</sub>Cl<sub>2</sub> solution. Remarkably, no premature cleavage of the oxime linker was observed at this higher concentration of acid. Attachment of the first pyrrole unit results in a bulky aromatic oxime ester that apparently stabilizes the linker relative to the aliphatic oxime ester present when conventional amino acids are used.

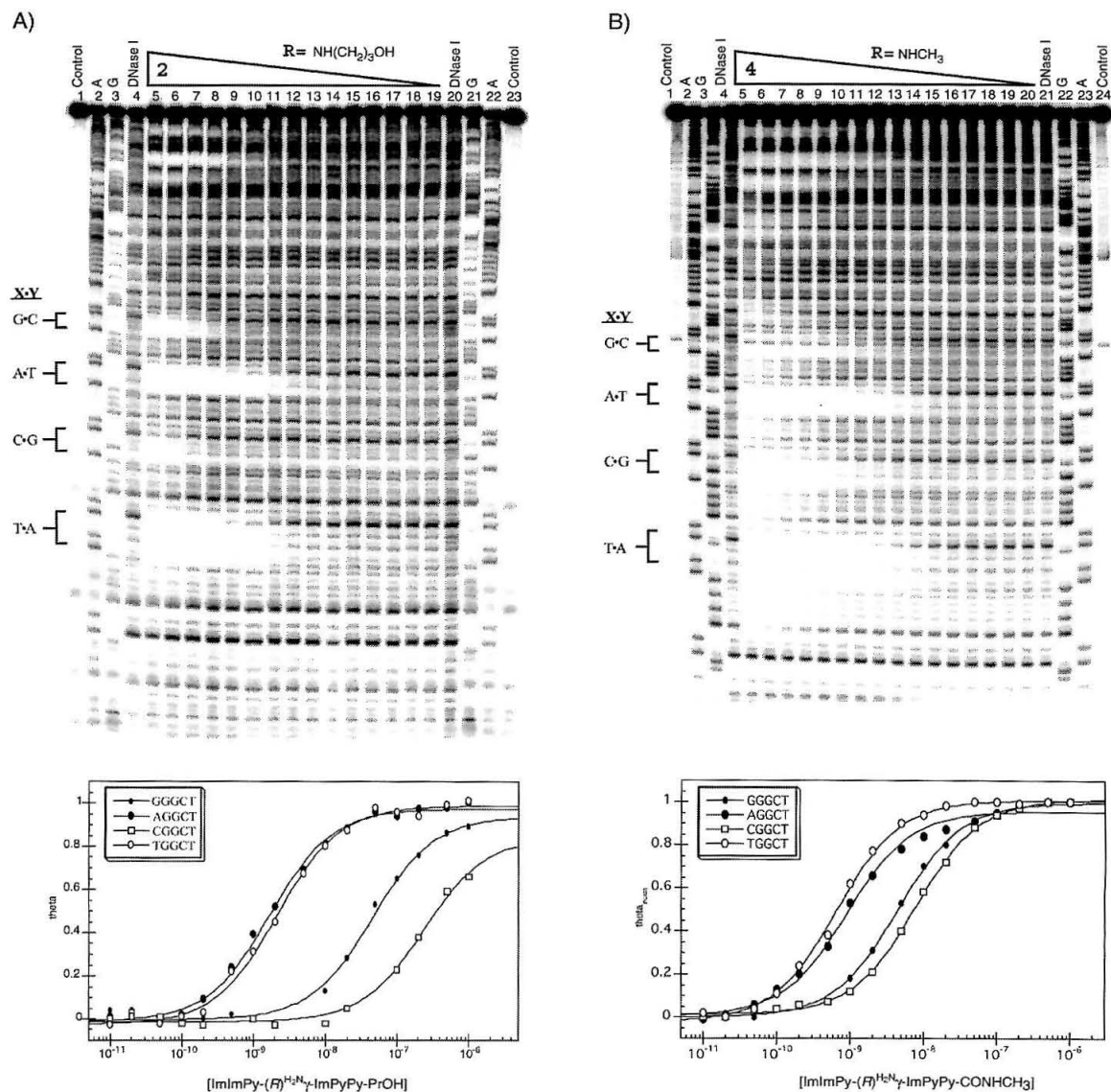
Polyamides **3-6** were generated from the common intermediate, ImImPy-(*R*)<sup>FmocHN</sup>γ-ImPyPy-Oxime resin **11** (Figure 5.3). Commercially available oxime resin was allowed to react directly with Boc-Py-OBt (**7**) in *N*-methylpyrrolidone (NMP) and diisopropylethylamine (DIEA) overnight at room temperature. The subsequent deprotections and couplings were carried out in a stepwise manner using the previously outlined deprotection conditions. All other monomers used, namely **8**, **9** and **10**, were activated with HBTU and coupled in NMP/DIEA for 1.5-8 hours at either room temperature or 37° C. All cleavages of resin **11** required a co-solvent (CH<sub>2</sub>Cl<sub>2</sub> or DMF) to swell the resin and/or solubilize the polyamide. Based on literature methods,<sup>10</sup> cleavage of resin **11** with a saturated solution of ammonia in THF/CH<sub>2</sub>Cl<sub>2</sub> yielded polyamide **5**. However, even after more than 60 hours at 37° C only 60% cleavage was observed. Addition of a large excess of DBU, which had been used as a transacylation catalyst to generate peptide acids and esters from oxime resin previously,<sup>13</sup> resulted in quantitative cleavage of the resin to generate polyamide **5**. The C-terminal carboxylic acid polyamide **6** was generated in a similar manner with 1:1 H<sub>2</sub>O:DMF, 30 eq. DBU for 60 hours at 37° C. This method was superior to acidic, basic, or reductive cleavage conditions. It is interesting to note that the standard transacylation catalyst, DMAP, was ineffective as a replacement for DBU in these cleavage reactions. Both polyamides **3** and **4** were generated from resin **11** with a 1.0 M solution in THF/CH<sub>2</sub>Cl<sub>2</sub> of ethylamine or methylamine, respectively overnight at 37° C. The oxime resin is a versatile way to



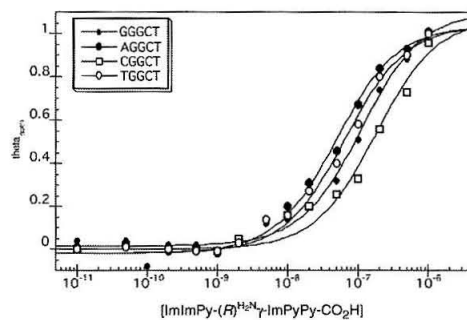
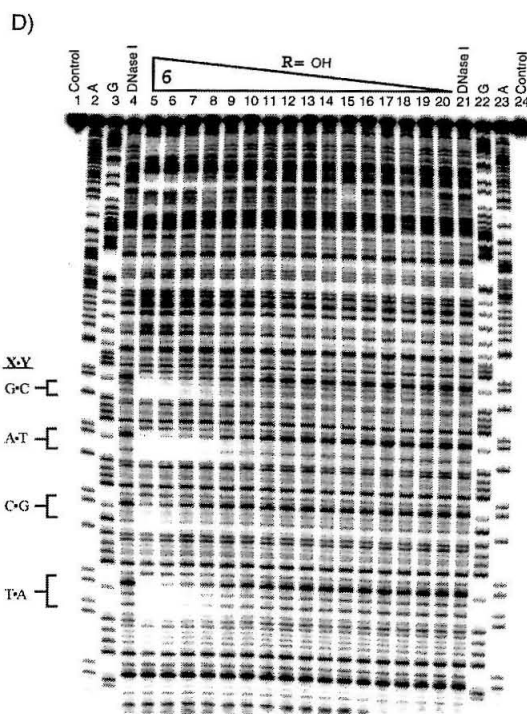
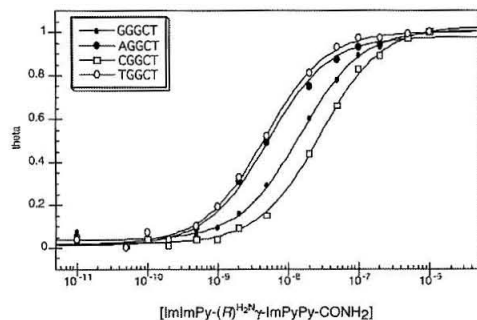
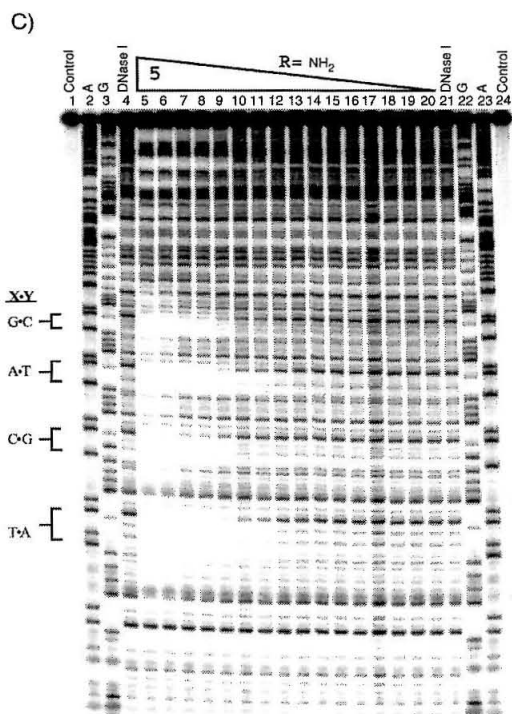
generate modified polyamides under mild conditions in good yield while remaining amenable to Boc chemistry.

**Quantitative DNase I Footprinting.** It has been shown previously that elimination of Dp in the  $\beta$ -Dp tail in polyamide **1** diminishes the requirement for A,T preference at the N-2 position.<sup>4</sup> While empirical observations have suggested that the propanol amide tail, like the  $\beta$ -Dp tail (**2**) is A,T specific at the N-1 position, this had not been investigated in detail. Quantitative DNase I footprinting titrations<sup>15</sup> were performed to determine the equilibrium association constants ( $K_a$ ) of polyamides **2-6** for four sites on a restriction fragment corresponding to the sequences 5'-AXGGCTA-3' (where X = A,T,G,C) at the N-1 position (Figure 4A-D, Table 1). The change from  $\beta$ -Dp (**1**) to propanol amide (**2**) results in similar affinity and specificity. This retention of affinity upon removal of the positively charged Dp moiety has previously been observed.<sup>16,17</sup> Presumably, the steric clash of propanol amide versus  $\beta$ -Dp with the exocyclic  $\text{NH}_2$  of G at N-1 is similar.

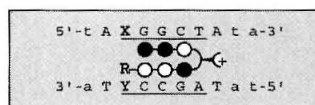
Polyamides **3-5** afford similar DNA binding affinity versus the parent  $\beta$ -Dp while diminishing the sequence preference for A,T versus G,C at N-1. Polyamide **3** (ethyl amide) maintains a strong preference for A,T versus G,C, while polyamide **4** (methyl amide) less so. Polyamide **5** (primary amide) shows a loss in affinity relative to **1-4**, and little sequence discrimination between the four Watson-Crick base pairs. Polyamide **6** (carboxylic acid) reveals poor affinity and little sequence discrimination at the N-1 position (Table 5.1). The decreased affinity of polyamides **5** and **6** suggests that the increased Van der Waal contacts with the walls of the minor groove, made by polyamides **1-4**, may be important.

**Figure 5.4**

**Figure 5.4.** All reactions contain 15 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. Data was obtained for the four binding sites indicated at the left of the gel, 5'-GGGCT-3', 5'-AGGCT-3', 5'-CGGCT-3', and 5'-TGGCT-3', and is shown in the isotherm plot below.  $\theta_{\text{norm}}$  points were obtained using storage phosphor autoradiography and processed by standard methods. Each data point shows the average value obtained from three footprinting experiments. The solid curves are best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm where  $n = 1$  as previously described.<sup>15</sup> **A)** Quantitative DNase I footprint titration experiment with ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyPrOH (**2**) on the 3'-<sup>32</sup>P-labeled 286-bp restriction fragment pSES-TL1<sup>4</sup>: lane 1 and 23, intact DNA; lane 2 and 22, A reaction; lane 3 and 21, G reaction; lane 4 and 20 DNase I standard; lanes 5-19 DNase I digestion products in the presence of 1 μM, 500 nM, 200 nM, 100 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide, respectively. **B)** Quantitative DNase I footprint titration experiment with ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONHCH<sub>3</sub> (**4**) on the 3'-<sup>32</sup>P-labeled 286-bp restriction fragment pSES-TL1<sup>4</sup>: lane 1 and 24, intact DNA; lane 2 and 23, A reaction; lane 3 and 22, G reaction; lane 4 and 21 DNase I standard; lanes 5-20 DNase I digestion products in the presence of 1 μM,



500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide, respectively. C) Quantitative DNase I footprint titration experiment with ImImPy-(R)H<sub>2</sub>N-ImPyPyCONH<sub>2</sub> (5) on the 3'-<sup>32</sup>P-labeled 286-bp restriction fragment pSES-TL1<sup>4</sup>: lane 1 and 24, intact DNA; lane 2 and 23, A reaction; lane 3 and 22, G reaction; lane 4 and 21 DNase I standard; lanes 5-20 DNase I digestion products in the presence of 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide, respectively. D) Quantitative DNase I footprint titration experiment with ImImPy-(R)H<sub>2</sub>N-ImPyPyCO<sub>2</sub>H (6) on the 3'-<sup>32</sup>P-labeled 286-bp restriction fragment pSES-TL1<sup>4</sup>: lane 1 and 24, intact DNA; lane 2 and 23, A reaction; lane 3 and 22, G reaction; lane 4 and 21 DNase I standard; lanes 5-20 DNase I digestion products in the presence of 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide, respectively.



**Table 5.1.** Equilibrium association constants ( $M^{-1}$ ).

	R= $-\beta$ -Dp (1)	$-\text{NH}(\text{CH}_2)_3\text{OH}$ (2)	$-\text{NHCH}_2\text{CH}_3$ (3)	$-\text{NHCH}_3$ (4)	$-\text{NH}_2$ (5)	$-\text{OH}$ (6)
X•Y						
A•T	$1.7 \times 10^9(\pm 0.5)$	$5.9 \times 10^8(\pm 0.4)$	$1.1 \times 10^9(\pm 0.1)$	$1.1 \times 10^9(\pm 0.1)$	$1.9 \times 10^8(\pm 0.2)$	$2.1 \times 10^7(\pm 0.1)$
T•A	$1.6 \times 10^9(\pm 0.2)$	$4.7 \times 10^8(\pm 0.4)$	$9.2 \times 10^8(\pm 0.3)$	$1.4 \times 10^9(\pm 0.3)$	$2.1 \times 10^8(\pm 0.2)$	$1.2 \times 10^7(\pm 0.2)$
G•C	$5.8 \times 10^7(\pm 1.4)$	$2.1 \times 10^7(\pm 0.3)$	$8.2 \times 10^7(\pm 1.3)$	$2.4 \times 10^8(\pm 0.1)$	$6.8 \times 10^7(\pm 0.5)$	$1.1 \times 10^7(\pm 0.2)$
C•G	$3.2 \times 10^7(\pm 0.3)$	$6.2 \times 10^6(\pm 0.7)$	$3.0 \times 10^7(\pm 0.5)$	$1.1 \times 10^8(\pm 0.2)$	$3.5 \times 10^7(\pm 0.2)$	$4.9 \times 10^6(\pm 0.2)$

**Table 5.1.** Polyamide 1-6 at the N-1 position. Equilibrium association constants ( $M^{-1}$ ) for polyamide 1-6 with restriction fragment pSES-TL1.<sup>4</sup> Values reported are the mean values from at least three DNase I footprinting titration experiments with the standard deviation for each data set in parentheses. Equilibrium association constants for polyamide 1 at the N-1 position were obtained from Swalley *et al.*<sup>4</sup> Assays were performed at 22° C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM  $\text{MgCl}_2$ , and 5 mM  $\text{CaCl}_2$ .

## Discussion

In sequences with an A•T or T•A at the N-1 position, structural evidence<sup>18-22</sup> suggests that the C-terminal pyrrole secondary amide makes a hydrogen bond with the N3 of adenine and/or the O2 of thymine (Figure 1). However, hydrogen bonds could also be made to the N3 of guanine and/or the O2 of cytosine. Assuming these hydrogen bonds are similar energetically, the discrimination of A,T versus G,C is likely due to steric reasons. Thus, the methylene groups of the propanol amide and  $\beta$ -Dp tails clash with the exocyclic amine of guanine and by default incur a lower energetic penalty at A,T. The influence of the exocyclic amine of guanine on specificity can be inferred from the observation that the sequence XY = C•G was the lowest affinity site for all of the polyamides examined in this study. At this site the guanine base is on the bottom strand of the DNA proximal to the C-terminus of the hairpin where the steric clash between the tail and the exocyclic  $\text{NH}_2$  of guanine should be most severe.

Polyamides **3** and **4** with the ethyl and methyl amide tails provide a way to investigate whether specificity could be assigned to a particular methylene group of the propanol amide tail. The specific hydrogen bond made by the amide of the C-terminal pyrrole would point the first methylene group away from the floor of the minor groove, such that the energetic penalty and hence specificity might be attributable to the second methylene group. However, the methyl amide did exhibit about 8-fold preference for binding A,T or G,C. The ethyl amide tail disfavors both G•C and C•G sites by about 3-fold more than the methyl amide tail. Clearly, both alkyl groups contribute to A,T specificity, and even greater steric bulk beyond the second carbon is likely necessary to account for the greater than 50-fold specificity of the  $\beta$ -Dp and propanol amide tails. The high affinity of polyamides **3** and **4** suggests an energetic balance between increased Van der Waals contacts with the walls of the minor groove relative to the primary amide tail (**5**) and decreased steric clashing relative to the  $\beta$ -Dp tail (**1**).

Polyamides with C-terminal carboxylic acid tails have been synthesized as intermediates to cyclic polyamides<sup>23,24</sup> but their DNA binding properties have not been investigated. The carboxylic acid tail polyamide (**6**) binds all four sites with low affinity and poor sequence discrimination. To act as a hydrogen bond donor in analogy to the amides, the carboxylic acid would have to be protonated in the minor groove. The acid is likely deprotonated in the minor groove and the 10-fold lower affinity compared to the primary amide tail may be due to loss of this hydrogen bond. Additionally, in water polyamide **6** is expected to be a zwitterion with a net charge of zero which one would expect to have a lower affinity for DNA than the positively charged polyamides **1-5**.

In summary, hairpin polyamides with incrementally truncated C-terminal tails can be synthesized using Boc chemistry, in high yield, on an oxime resin by solid-phase methods. The polyamides are lower molecular weight than the corresponding  $\beta$ -Dp polyamide and bind all four base pairs at the N-1 position with high affinity which will enable targeting of new biologically relevant sequences by hairpin polyamides. If at the

N-1 position, one desires specificity for A,T > G,C while maintaining high affinity, then the ethyl amide (**3**) is the end group of choice. If at the N-1 position, one prefers less sequence preference with high affinity, then the methyl amide is the best choice. It remains to be seen whether we can design an energetically favorable end group compatible with solid-phase methods with the criteria of efficient cleavage from solid-phase resin which favors G,C > A,T.

## Experimental

**Materials.** Polyamides **1** and **2**, and the restriction fragment pSES-TL1 have been previously described.<sup>4</sup> DNase I footprinting titrations on the <sup>32</sup>P-labeled restriction fragment pSES-TL1 were performed as previously described.<sup>4</sup> 2.0 M methylamine in THF, 2.0 M ethylamine in THF, DBU and anhydrous ammonia gas were purchased from Aldrich. Oxime resin was purchased from Nova Biochem. All other synthetic and footprinting reagents were as previously described.<sup>2,15,16</sup> <sup>1</sup>H NMR were recorded on a Varian Mercury 300 instrument. UV spectra were measured on a Beckman Coulter DU 7400 diode array spectrophotometer. Autoradiography was performed with a Molecular Dynamics Typhoon Phosphorimager. Matrix-assisted, laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry was carried out at the Peptide and Protein Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold system using a RAINEN C<sub>18</sub>, Microsorb MV, 5 µm, 300 x 4.6 mm reversed-phase column in 0.1% (w/v) TFA with CH<sub>3</sub>CN as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% CH<sub>3</sub>CN/min. Preparatory HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100 µm C<sub>18</sub> column, 0.1% (w/v) TFA, 0.56% CH<sub>3</sub>CN/min. 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 µm filtered. Reagent-grade chemicals were used unless otherwise stated.



**ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONHEt (3).** ImImPy-(R)<sup>FmocHN</sup>γ-ImPyPyCO-Oxime resin (**11**) was generated by manual solid phase synthesis from oxime resin (1 g, 0.48 mmol/g) using previously described Boc-protected monomers.<sup>2</sup> Boc-Py-OBt (**7**) (358 mg, 1 mmol) was dissolved in 2 ml of NMP and added to 1 g of oxime resin followed by 1 ml of DIEA. The coupling was allowed to proceed overnight at room temperature. The resin was then acetylated with 3 ml of acetic anhydride (Ac<sub>2</sub>O), 4 ml of NMP and 1 ml of DIEA for 30 minutes at room temperature. The Boc group was removed upon treatment with 20% TFA/DCM for 30 minutes. The second pyrrole residue was coupled in the same fashion as the first but complete coupling could be achieved in 2 hours at room temperature followed by the acetylation and deprotection steps outlined above. Boc-Im-OH (**8**) (241 mg, 1 mmol) was dissolved in 2 ml of NMP to which 360 mg (1 mmol) of HBTU and 1 ml of DIEA was added for activation of this monomer. Coupling was allowed to proceed for 1.5 hours at room temperature followed by acetylation. The Boc-Im residue was deprotected using a 50% TFA/DCM solution for 30 minutes at room temperature. The next residue, α-Fmoc-γ-Boc-(R)-diaminobutyric acid (DABA) (**9**) (660 mg, 1.5 mmol) was activated with HBTU (540 mg, 1.5 mmol) in 2 ml of NMP and 1 ml of DIEA. Coupling of this residue onto the resin took 2 hours at 37° C. After acetylation and treatment with 20% TFA/DCM for 30 minutes, the next Boc-Py-OBt was attached in exactly the same manner as the second residue. The terminal two imidazoles were added as an ImIm-OH dimer.<sup>2</sup> 249 mg (1 mmol) of the ImIm-OH dimer (**10**) was activated with HBTU (360 mg, 1 mmol) in 2 ml of NMP and 1 ml of DIEA and allowed to couple overnight at room temperature. It should be noted that the progress of the stepwise couplings were all monitored by analytical HPLC. The resin was washed thoroughly with DMF, DCM, MeOH and Et<sub>2</sub>O then dried *in vacuo*. A 75 mg sample of dried resin was suspended in 2 ml of CH<sub>2</sub>Cl<sub>2</sub> to which was added 2 ml of 2.0 M ethylamine in THF. This cleavage mixture was placed in a 37° C oven and allowed to stand overnight in a sealed scintillation vial. The resin was filtered, the eluant concentrated *in vacuo*, then



purified by reverse phase HPLC. ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONHCH<sub>3</sub> (1.9 mg, 2.2 μmol, 10.6% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H<sub>2</sub>O) λ<sub>max</sub> 310 (52140); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>) δ 11.02 (s, 1H), 10.36 (s, 1H), 10.10 (s, 1H), 9.89 (s, 1H), 9.71 (s, 1H), 8.20 (m, 1H), 7.98 (m, 1H), 7.56 (s, 1H), 7.52 (s, 1H), 7.45 (d, 1H, *J* = 0.9 Hz), 7.25 (d, 1H, *J* = 1.5 Hz), 7.23 (d, 1H, *J* = 1.5 Hz), 7.15 (d, 1H, *J* = 1.8 Hz), 7.11 (d, 1H, *J* = 1.8 Hz), 7.06 (d, 1H, *J* = 0.9 Hz), 7.03 (d, 1H, *J* = 1.8 Hz), 6.83 (d, 1H, *J* = 1.8 Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.83 (s, 3H), 3.80 (s, 3H), 3.78 (s, 3H), 3.17 (dd, 2H, *J* = 6.6 Hz), 2.42 (m, 2H), 2.26 (m, 1H), 1.99 (m, 2H), 1.62 (m, 1H), 1.05 (t, 3H *J* = 6.6 Hz), MALDI-TOF-MS 866.5 (866.39 calc for [M+H] C<sub>39</sub>H<sub>48</sub>N<sub>17</sub>O<sub>7</sub><sup>+</sup>).

**ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONHMe (4).** A 35 mg sample of dried resin **11** was suspended in 2 ml of CH<sub>2</sub>Cl<sub>2</sub> to which was added 2 ml of 2.0 M methylamine in THF. This cleavage mixture was placed in a 37° C oven and allowed to stand overnight in a sealed scintillation vial. The resin was filtered, the eluant concentrated *in vacuo*, then purified by reverse phase HPLC. ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONHCH<sub>3</sub> (850 μg, 996 nmol, 8.9% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H<sub>2</sub>O) λ<sub>max</sub> 310 (52140); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>) δ 11.01 (s, 1H), 10.35 (s, 1H), 10.09 (s, 1H), 9.90 (s, 1H), 9.70 (s, 1H), 8.19 (m, 1H), 7.92 (m, 1H), 7.56 (s, 1H), 7.51 (s, 1H), 7.45 (d, 1H, *J* = 0.9 Hz), 7.25 (d, 1H, *J* = 1.8 Hz), 7.23 (d, 1H, *J* = 1.8 Hz), 7.16 (d, 1H, *J* = 1.5 Hz), 7.10 (d, 1H, *J* = 1.5 Hz), 7.05 (d, 1H, *J* = 0.9 Hz), 7.03 (d, 1H, *J* = 1.8 Hz), 6.79 (d, 1H, *J* = 1.8 Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.82 (s, 3H), 3.79 (s, 3H), 3.77 (s, 3H), 2.66 (m, 1H), 2.41 (m, 2H), 2.25 (m, 1H), 1.98 (m, 2H), 1.62 (m, 2H), MALDI-TOF-MS 852.5 (852.38 calc for [M+H] C<sub>38</sub>H<sub>46</sub>N<sub>17</sub>O<sub>7</sub><sup>+</sup>).

**ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONH<sub>2</sub> (5).** A 30 mg sample of dried resin **11** was placed into a pressure tolerant screw cap test tube, suspended in 4 ml of dry THF and

cooled for 20 minutes in a  $-10^{\circ}\text{C}$  ice/brine bath. Anhydrous  $\text{NH}_3$  gas was bubble into the suspension at a steady rate for 30 minutes to generate a saturated ammonia solution. 1.5 ml of a 10%v/v solution of DBU in  $\text{CH}_2\text{Cl}_2$ , cooled to  $-20^{\circ}\text{C}$  was quickly added to the cleavage mixture and the tube immediately sealed. The cleavage was allowed to proceed in a shaker at  $37^{\circ}\text{C}$  for 72 hours. The mixture is filtered and concentrated *in vacuo*, then purified by reverse phase HPLC. ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONH<sub>2</sub> (330 μg, 390 nmol, 4.1% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV ( $\text{H}_2\text{O}$ )  $\lambda_{\text{max}}$  310 (52140);  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  10.65 (s, 1H), 10.32 (s, 1H), 10.20 (s, 1H), 9.90 (s, 1H), 9.75 (s, 1H), 8.09 (bs, 1H), 7.55 (s, 3H), 7.47 (s, 1H), 7.44 (s, 1H), 7.25 (d, 1H,  $J = 2.1$  Hz), 7.21 (d, 1H,  $J = 2.1$  Hz), 7.19 (bs 2H), 7.11 (d, 1H,  $J = 1.8$  Hz), 7.05 (d, 1H,  $J = 1.2$  Hz), 6.98 (d, 1H,  $J = 1.2$  Hz), 6.83 (d, 1H,  $J = 1.5$  Hz), 6.65 (d, 1H,  $J = 1.2$  Hz), 6.53 (d, 1H,  $J = 2.4$  Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.82 (s, 3H), 3.79 (s, 1H), 3.77 (s, 3H), 2.97 (m, 1H), 1.96 (m, 2H), 1.43 (m, 2H), MALDI-TOF-MS 838.4 (838.35 calc for  $[\text{M}+\text{H}] \text{C}_{37}\text{H}_{44}\text{N}_{17}\text{O}_7^+$ )

**ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCO<sub>2</sub>H (6).** A 35 mg sample of dried resin **11** was cleaved by treatment with 1 ml of  $\text{H}_2\text{O}$ , 1 ml DMF, and 75 μl DBU (0.48 mmol) at  $37^{\circ}\text{C}$  for 60 hours. After filtration the cleavage solution was concentrated *in vacuo* and purified by reverse phase HPLC. ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCO<sub>2</sub>H (294 μg, 350 nmol, 3.1% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV ( $\text{H}_2\text{O}$ )  $\lambda_{\text{max}}$  310 (52140);  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  12.67 (bs, 1H), 11.53 (s, 1H), 10.88 (s, 1H), 10.62 (s, 1H), 10.43 (s, 1H), 10.22 (s, 1H), 8.83 (m, 1H), 8.07 (s, 1H), 8.03 (s, 1H), 7.96 (d, 1H,  $J = 0.9$  Hz), 7.93 (d, 1H,  $J = 1.8$  Hz), 7.76 (d, 1H,  $J = 0.9$  Hz), 7.75 (d, 1H,  $J = 1.8$  Hz), 7.64 (d, 1H,  $J = 1.2$  Hz), 7.57 (d, 1H,  $J = 1.2$  Hz), 7.54 (d, 1H,  $J = 1.8$  Hz), 7.34 (d, 1H,  $J = 1.8$  Hz), 4.50 (s, 3H), 4.50 (s, 3H), 4.47 (s, 3H), 4.34 (s, 3H), 4.31 (s, 3H), 4.31 (s, 3H), 2.50 (m, 2H), 2.11 (m, 2H), 2.06 (m, 3H), MALDI-TOF-MS 839.4 (839.34 calc for  $[\text{M}+\text{H}] \text{C}_{37}\text{H}_{43}\text{N}_{16}\text{O}_8^+$ ).

**Quantitative DNase I Footprinting.** As previously reported.<sup>15</sup>

## Acknowledgement

We are grateful to the National Institutes of Health GM27681 for research support, the Ralph M. Parsons Foundation for a predoctoral fellowship to J.M.B., the Natural Sciences and Engineering Research Council of Canada for a postgraduate scholarship to D.H.N., and Bristol-Myers Squibb for a predoctoral fellowship to N.R.W.

## References

- (1) Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688-693.
- (2) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
- (3) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147-6152.
- (4) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.
- (5) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chem.-Int.Edit. Engl.* **1998**, *37*, 1421-1423.
- (6) Gottesfeld, J. M.; Turner, J. M.; Dervan, P. B. *Gene Expression* **2000**, *9*, 77-91.
- (7) Ioshikhes, I. P.; Zhang, M. Q. *Nature Genetics* **2000**, *26*, 61-63.
- (8) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053-1095.
- (9) DeGrado, W. F., Kaiser, E.T. *J. Org. Chem.* **1980**, *45*, 1295-1300.
- (10) DeGrado, W. F., Kaiser, E.T. *J. Org. Chem.* **1982**, *47*, 3258-3261.
- (11) Behrens, C., Nielson, P.E. *Combinatorial Chemistry & High Throughput Screening* **1998**, *1*, 127-134.
- (12) Voyer, N., Lavoie, A., Pinette, M., Bernier, J. *Tetrahedron Lett.* **1994**, *35*, 355-358.

- (13) Pichette, A., Voyer, N., Larouche, R., Meillon, J-C. *Tetrahedron Lett.* **1997**, 38, 1279-1282.
- (14) Nakagawa, S. H., Kaiser, E.T. *J. Org. Chem.* **1983**, 48, 678-685.
- (15) Trauger, J. W., Dervan, P.B. *Methods Enzymol.* **2001**, 340, 450-466.
- (16) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, 120, 1382-1391.
- (17) Bremer, R. E., Wurtz, N.R., Szewczyk, J.W., Dervan, P.B. *Bioorg. Med. Chem.* **2001**, 9, 2093-2103.
- (18) deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, 119, 7909-7916.
- (19) deClairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1999**, 121, 2956-2964.
- (20) Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C. *Nat. Struct. Biol.* **1998**, 5, 104-109.
- (21) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, 295, 557-567.
- (22) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, 282, 111-115.
- (23) Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 10389-10392.
- (24) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, 121, 1121-1129.

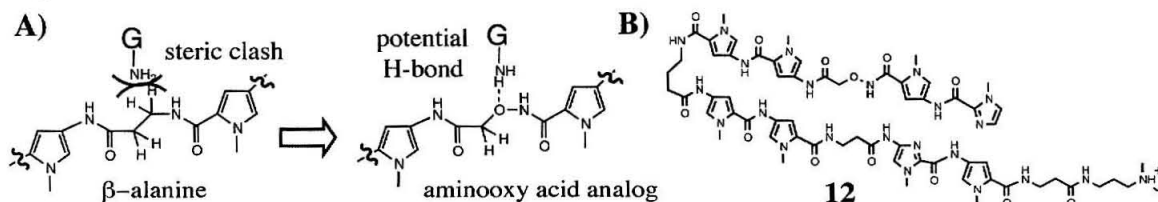
## Chapter 5B:

### From an Aminooxy $\beta$ -Alanine Analog to Novel Polyamide Tails<sup>1</sup>

#### Introduction

The development of polyamide synthesis on the versatile oxime resin allows for a large variety of C-termini, including the “truncated tails” described in Chapter 5A. The primary amide (**5**) was the first of this class of compounds synthesized, prior to the development of polyamide synthesis on oxime resin. The initial route began from the unusual starting point of an aminooxy acid. Aminooxy acids are analogs of  $\beta$ -amino acids where the  $\beta$ -carbon is replaced by oxygen. Peptides of aminooxy acids, which have hydroxamic ester linkages instead of normal amide bonds, have attracted attention as peptidomimetics.<sup>2-6</sup> In the context of DNA binding polyamides, an aminooxy acid was investigated as a potential G-selective  $\beta$ -alanine analog (Figure 5.5).<sup>7</sup> That project was ultimately unsuccessful due to unusual DNA binding properties of the resulting polyamide.<sup>8</sup> However, it was successful synthetically, in that conditions were found for the incorporation of an aminooxy acid unit into a full-length polyamide.

**Figure 5.5**



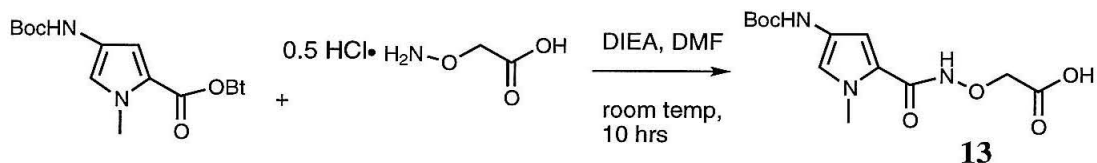
**Figure 5.5.** A) Model of an aminooxy acid as a potential G-selective  $\beta$ -alanine analog.<sup>7</sup> B) The polyamide (**12**) synthesized to test the model. The aminooxy acid unit is shown in red.

Given the compatibility of the hydroxamic ester with standard solid phase synthesis conditions,<sup>9</sup> it was reasoned that the chemistry of the N-O bond could be exploited to produce polyamides with C-terminal pyrrole primary amides. The aminooxy

acid unit was incorporated as a linker between the resin and the growing polyamide chain. After completion of resin synthesis, the well-known reduction of hydroxamic esters to amides<sup>10,11</sup> was accomplished on solid phase,<sup>12,13</sup> cleaving the C-terminal primary amide polyamide from the resin. The advantage of this route is that the target compound is obtained using standard solid phase conditions for polyamide synthesis (Boc-protected monomers and PAM resin).<sup>9</sup> Subsequently a complimentary route to primary amide polyamides was devised, using recently developed Fmoc-protected monomers<sup>14</sup> and Rink resin.<sup>15</sup> Finally, polyamide synthesis was developed on the extremely versatile oxime resin (Chapter 5A), providing access to a variety of truncated tails, including the primary amide. However, solid phase synthesis with the aminooxy linker also yielded a variety of different C-termini, due to an unexpected reaction with cyclic secondary amines. The resulting polyamides with C-terminal cyclic tertiary amides have favorable DNA binding properties based on preliminary Dnase I footprinting titrations.

## Results

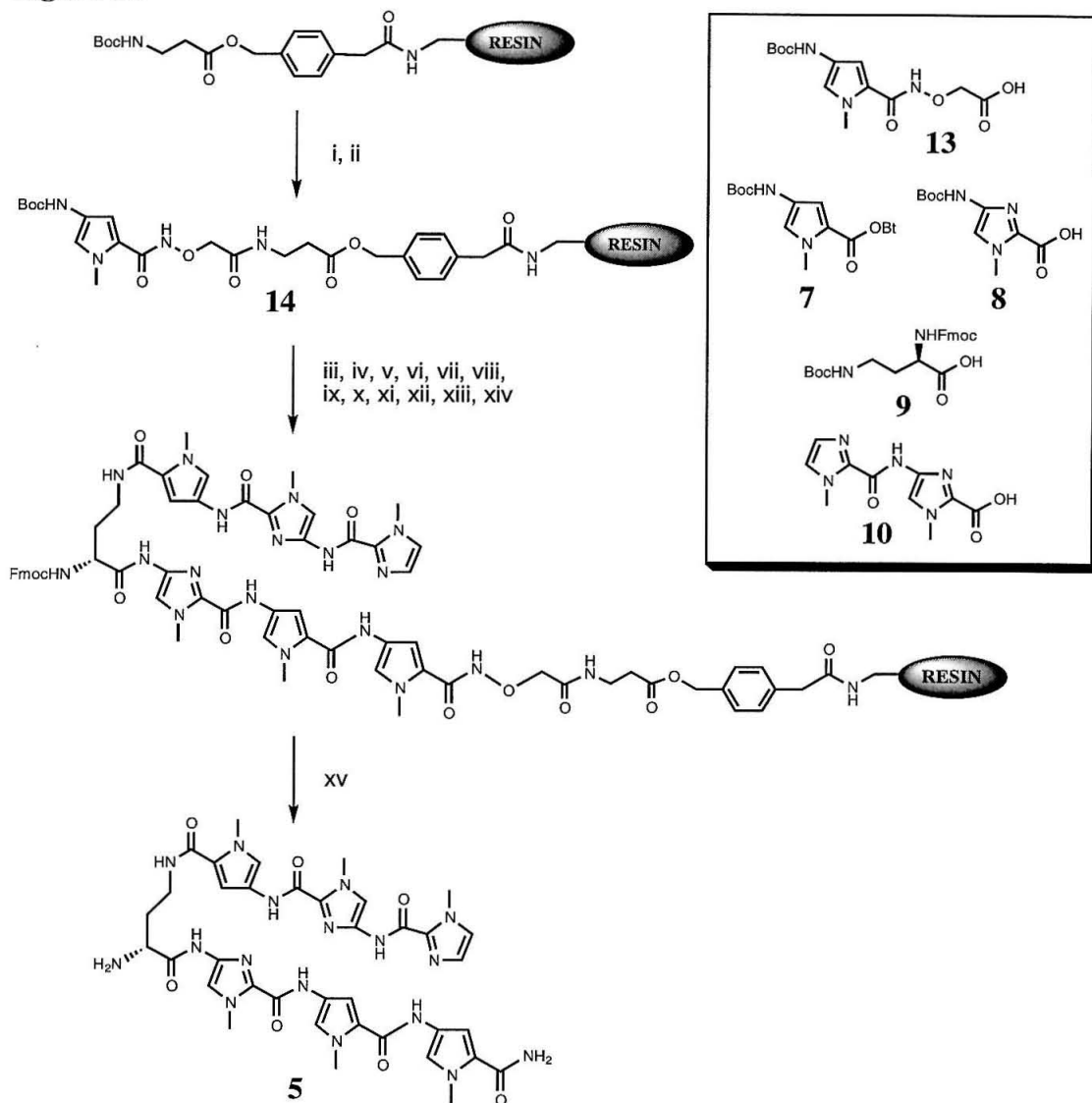
**Aminooxy linker route to C-terminal primary amide polyamides.** Although a Boc-protected aminooxy acid analog of  $\beta$ -alanine is readily accessible in one step from commercially available carboxymethoxylamine hemihydrochloride,<sup>16</sup> use of this monomer under standard solid phase conditions<sup>9</sup> was highly problematic. Difficulties with primary aminooxy groups on solid phase have been previously noted.<sup>17,18</sup> After considerable experimentation, a simple solution presented itself in the one step, solution phase synthesis of the Boc-pyrrole-aminooxy acid (**13**) derived from the standard HOBt activated pyrrole monomer<sup>9</sup> (**7**) and carboxymethoxylamine hemihydrochloride (Figure 5.6). This unit was found to be compatible with standard solid phase conditions, and is readily incorporated into a polyamide chain by activation as the *N*-hydroxysuccinimide ester, followed by reaction with the amine of the pyrrole residue (as in **12**) or  $\beta$ -alanine.

**Figure 5.6****Figure 5.6.** Synthesis of Boc-pyrrole-aminooxy acid **13**.

The synthesis of **5** (Figure 5.7) began with the deprotection of commercially available Boc- $\beta$ -Ala-PAM resin, followed by coupling of *N*-hydroxysuccinimide (NHS) activated **13**, in DMF for six hours at room temperature with DIEA as a base. Removal of the Boc group from the resulting Boc-PyCONHOCH<sub>2</sub>CO- $\beta$ -Ala-PAM resin (**14**) was accomplished using standard deprotection conditions (80% TFA, 20% DCM, 0.4 M thiophenol). Five standard coupling and deprotection cycles<sup>9,19</sup> followed to yield ImImPy-(*R*)<sup>FmocHN</sup> $\gamma$ -ImPyPyCONHOCH<sub>2</sub>CO- $\beta$ -Ala-PAM resin. Some degradation was observed in a coupling which requires heating at 37° C, suggesting that the hydroxamic ester unit is best handled at room temperature, which may limit the generality of this method. Attempts to remove the Fmoc group with piperidine led to unexpected cleavage from the solid support (see below), therefore the Fmoc group was not removed prior to cleavage from resin. After surveying various reducing agents including zinc dust<sup>11</sup> and triphenylphosphine, best results for the reduction of the hydroxamic ester in the full length polyamide resin were obtained with palladium acetate and ammonium formate.<sup>20</sup> The Fmoc group was also removed under these reductive conditions<sup>21</sup> to yield the truncated tail polyamide, ImImPy-(*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ -ImPyPyCONH<sub>2</sub> (**5**). Results of quantitative DNase I footprinting titrations with **5** are reported in Chapter 5A (Figure 5.4C).

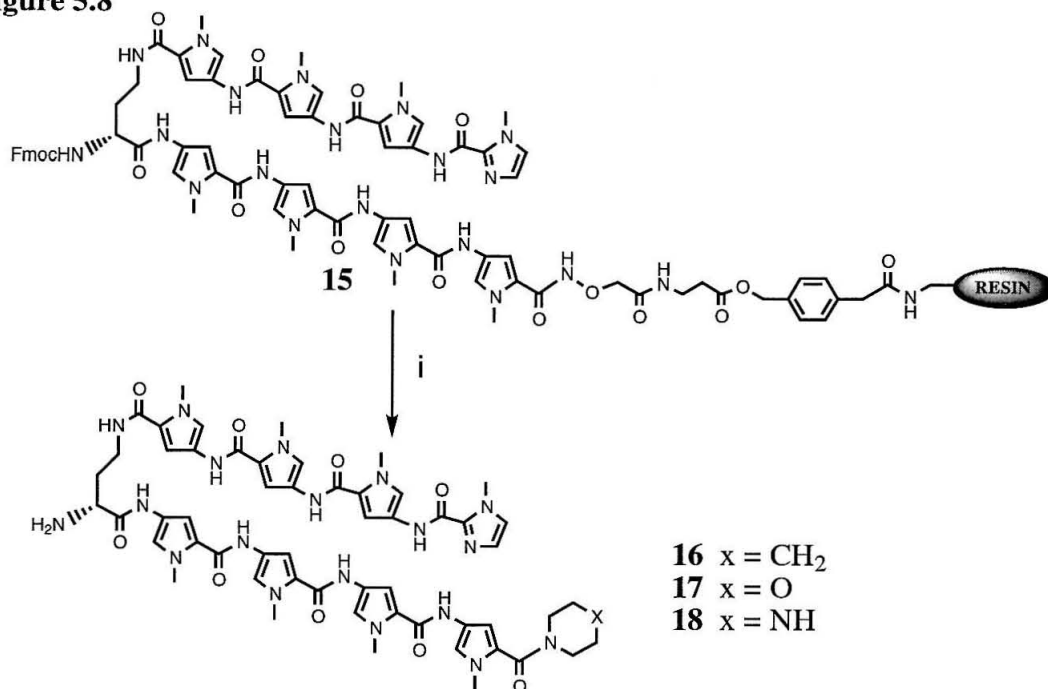
**Aminooxy linker route to C-terminal cyclic tertiary amide polyamides.** An eight ring polyamide resin, ImPyPyPy-(*R*)<sup>FmocHN</sup> $\gamma$ -PyPyPyPyCONHOCH<sub>2</sub>CO- $\beta$ -Ala-PAM resin (**15**) (Figure 5.8), was constructed in eight standard coupling and deprotection



**Figure 5.7**

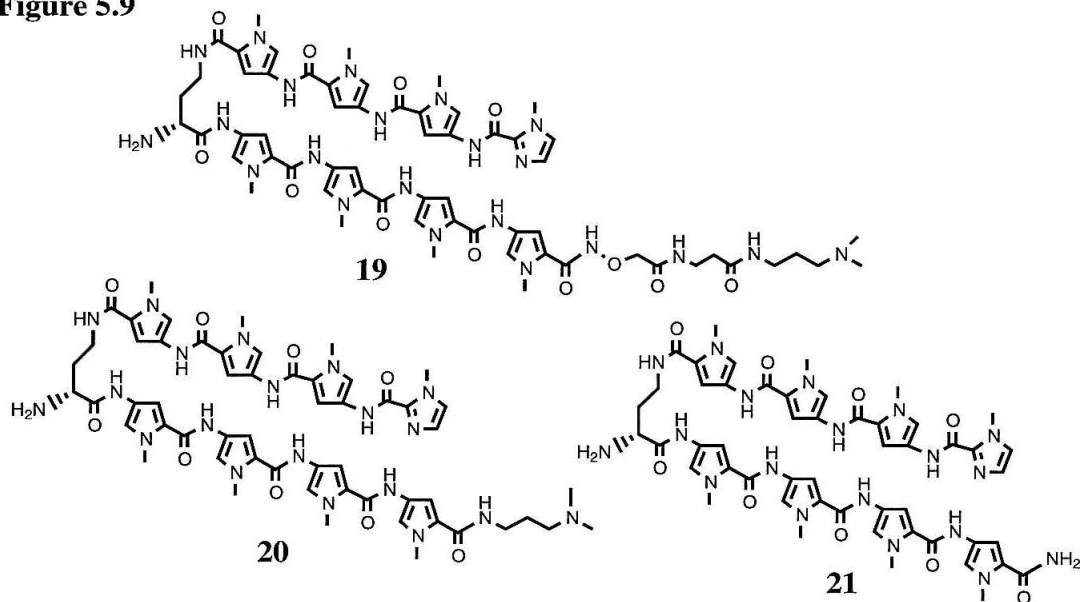
**Figure 5.7.** Aminooxy linker route to polyamide **5** starting from Boc- $\beta$ -Ala-PAM resin: (i) 80% TFA/DCM, 0.4 M PhSH; (ii) **13**, DCC, NHS, DIEA, DMF; (iii) 80% TFA/DCM, 0.4 M PhSH; (iv) **7**, DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) **8**, HBTU, DIEA, DMF, 37°C; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) **9**, HBTU, DIEA, DMF; (ix) 80% TFA/DCM, 0.4 M PhSH; (x) **7**, DIEA, DMF; (xi) 80% TFA/DCM, 0.4 M PhSH; (xii) **10**, HBTU, DIEA, DMF (xiii) Pd(OAc)<sub>2</sub>, HCO<sub>2</sub>NH<sub>4</sub>, H<sub>2</sub>O, DMF, 37°C.

cycles from resin **14**. Resin **15** was initially treated with 80% piperidine/DMF for 30 minutes at room temperature to remove the Fmoc group. Loss of material from the resin was visible by eye and reverse phase HPLC analysis. The putative deprotection reaction was repeated on a new sample of resin **15**, and a product was isolated from the resulting

**Figure 5.8**

**Figure 5.8.** Aminooxy linker route to C-terminal cyclic tertiary amide polyamides **16-18**. (i)  $x = \text{CH}_2$ , 2:1 piperidine:DMF, 1 hour, room temperature.  $x = \text{O}$ , 2:1 morpholine:DMF, 1 hour, room temperature.  $x = \text{NH}$ , 1 M piperazine in DMF, 1 hour, room temperature.

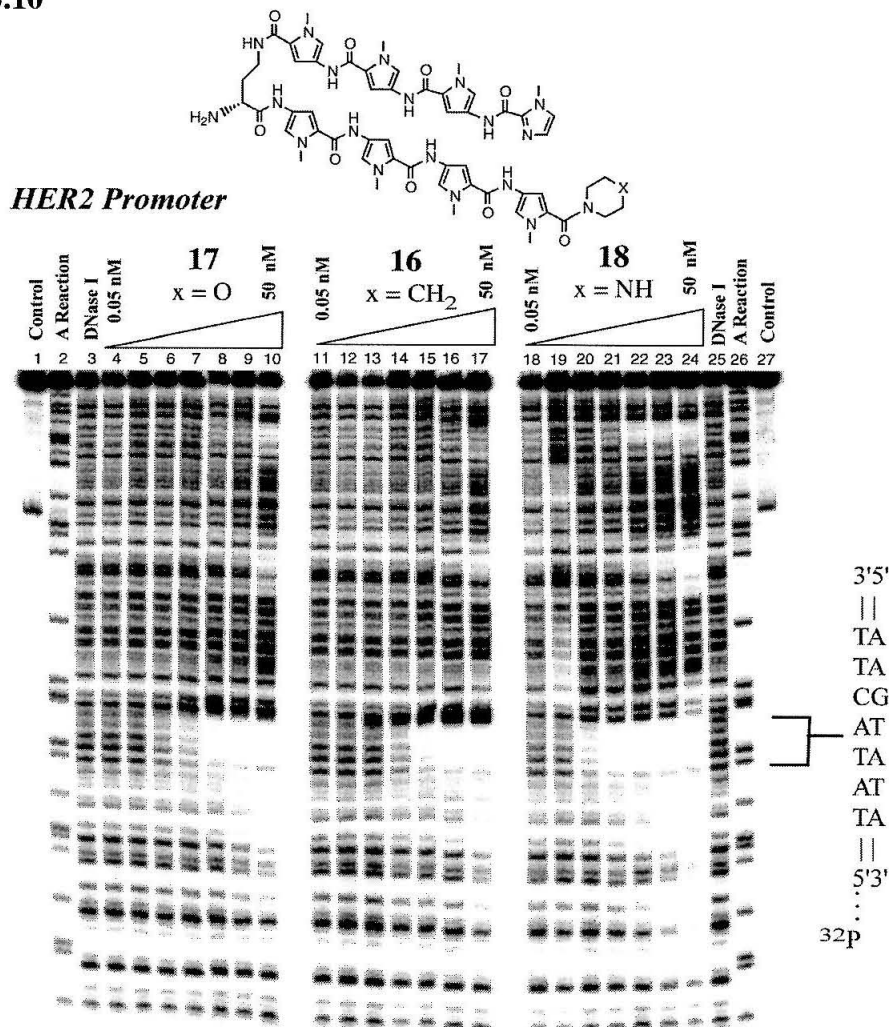
solution. The product was identified by HPLC, UV-vis, and MALDI-TOF mass spectrometry as being a relatively nonpolar polyamide with a mass consistent with the direct incorporation of piperidine as in structure **16** (Figure 5.8). Confirmation of the proposed structure was achieved synthetically by repeating the reaction with morpholine in place of piperidine. Indeed a more polar polyamide (**17**) with a molecular weight of two mass units higher was isolated. The unexpected cleavage reaction was also successful for piperazine, generating **18**, the most polar of the three new polyamides, with the predicted molecular weight. Optimized conditions for the synthesis of C-terminal cyclic tertiary amide polyamides **16-18** via the unexpected cleavage of resin **15** are shown in Figure 5.8. In all cases, the Fmoc group is also removed under these conditions. For polyamide-PAM resins lacking the aminooxy linker, deprotection of the Fmoc group is the only observed reaction. The reaction with resin **15**, which results in

**Figure 5.9**

**Figure 5.9.** Products obtained from the reaction of aminooxy linker resin **15** with Dp (**19** major, **20**, minor) or reductive cleavage conditions (**21**).

transamidation from a hydroxamic ester to a tertiary amide, occurs under remarkably mild conditions, one hour at room temperature with a large excess of the cyclic secondary amine. The reactions were quite clean, with **16-18** as the major pyrrole/imidazole-containing products, but the cleavage yield in all cases was below 50%, even with extended reaction time. In contrast to cyclic secondary amines in DMF, neat Dp reacted selectively with the PAM ester of resin **15** to yield the aminooxy acid containing polyamide **19** at both 37 °C (18 hours) and 95 °C (5-10 minutes), although the Dp transamidation product (**20**) was observed as a minor product (<10% of **19**) in the former case (Figure 5.9). Treatment of resin **15** with palladium acetate and ammonium formate led to expected primary amide polyamide **21**.

**Quantitative DNase I Footprinting.** Preliminary investigation of the C-terminal cyclic tertiary amide tail polyamides was carried out on a DNA fragment derived from the HER2 gene promoter.<sup>22</sup> Quantitative DNase I footprinting titrations<sup>23</sup> were performed

**Figure 5.10**

**Figure 5.4.** All reactions contain 15 kcpm DNA fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. Quantitative DNase I footprint titration experiment with C-terminal cyclic tertiary amide tail polyamides **16-18** on a 5'-<sup>32</sup>P-labeled 188-bp DNA fragment of the HER2 promoter<sup>22</sup>: lanes 1 and 27, intact DNA; lanes 2 and 26, A reaction; lanes 3 and 25 DNase I standard; lanes 4-10 DNase I digestion products in the presence of 50 pM, 100 pM, 1 nM, 5 nM, 10 nM, 50 nM polyamide **17**, respectively; lanes 11-17 DNase I digestion products in the presence of 50 pM, 100 pM, 1 nM, 5 nM, 10 nM, 50 nM polyamide **16**, respectively; lanes 18-24 DNase I digestion products in the presence of 50 pM, 100 pM, 1 nM, 5 nM, 10 nM, 50 nM polyamide **18**, respectively.

to determine the equilibrium association constants ( $K_a$ ) of polyamides **16-18** (Figure 5.10, Table 5.2). The polyamides bound the expected match site 5'-AAGTATA-3' (the TATA box of the HER2 promoter) with good specificity and affinity. The equilibrium association constants ( $7.5 \times 10^8 \text{ M}^{-1} - 5.0 \times 10^9 \text{ M}^{-1}$ ) are in the range typically observed for

**Table 5.2.** Equilibrium association constants.

Polyamide	Binding Site	$K_a$	DNA Fragment
<b>16</b>	5'-gAAGTATAag-3'	$7.5 \times 10^8 \text{ M}^{-1}$	HER2 Promoter
<b>17</b>	5'-gAAGTATAag-3'	$2.3 \times 10^9 \text{ M}^{-1}$	HER2 Promoter
<b>18</b>	5'-gAAGTATAag-3'	$5.0 \times 10^9 \text{ M}^{-1}$	HER2 Promoter
<b>17</b>	5'-tTAGTATTtg-3'	$2.4 \times 10^9 \text{ M}^{-1}$	pJT8
<b>22<sup>a</sup></b>	5'-tTAGTATTtg-3'	$3.5 \times 10^9 \text{ M}^{-1}$	pJT8
<b>17</b>	5'-tTAGTAC <u>T</u> tg-3'	$3.1 \times 10^8 \text{ M}^{-1}$	pJT8
<b>22<sup>a</sup></b>	5'-tTAGTAC <u>T</u> tg-3'	$5.0 \times 10^8 \text{ M}^{-1}$	pJT8

**Table 5.2.** Equilibrium association constants ( $\text{M}^{-1}$ ) for polyamides **16-18** with a 188-bp DNA fragment of the HER2 promoter,<sup>22</sup> and polyamides **17** and **22** with restriction fragment pJT8.<sup>24</sup> Mismatched base pairs are underlined. <sup>a</sup>Equilibrium association constants for polyamide **22** were obtained from Trauger *et al.*<sup>24</sup> Assays were performed at 22° C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

the standard  $\beta$ -Dp tail polyamide (ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp, **22**). The piperazine-amide polyamide **18** had the highest affinity, and the piperidine-amide polyamide **16** the least. Morpholine-amide polyamide **17**, which had the middle value on the HER2 fragment, was compared directly with standard polyamide **22** on the restriction fragment pJT8,<sup>24</sup> which contains the match site 5'-TAGTATT-3' and the mismatch site 5'-TAGTACT-3' (Table 5.2). The affinity and specificity observed for polyamide **17** was strikingly similar to the published values for **22**.<sup>24</sup>

## Discussion

In the course of developing an aminooxy linker route to C-terminal primary amide polyamides (**5** and **21**) an unexpected reaction was uncovered, which resulted in cleavage of the polyamide from the resin and replacement of the aminooxy linker with a cyclic secondary amine. The direct transamidation of a hydroxamic ester to a tertiary amide is unprecedented in the literature. Intermolecular transamidations usually take place only at

highly elevated temperatures, and very few transamidations of hydroxamic esters at any temperature are known.<sup>25-29</sup> However, the primary aminooxy group may be a better leaving group than it is generally considered.<sup>30</sup> For example, Canne *et al.* found that a 2-chlorobezilyoxy (2-Cl-Z) protecting group on an aminooxy acetyl group was unexpectedly sensitive to nucleophilic attack, as the 2-Cl-Z group was deprotected by mercaptoethanol.<sup>31</sup> The closest reaction in the literature to the resin cleavage observed here is the ring-opening of diacetyl-D-cycloserine, in which a highly activated acetylated hydroxamic lactam reacts very rapidly with n-butylamine in water at room temperature to give the ring-opened butylamide.<sup>29</sup> However, note that a large excess of the primary amine Dp at 37 °C or 95 °C was far less effective in the present case than a large excess of cyclic secondary amine in DMF at room temperature. Preliminary studies with model substrates such as **13** have not shown evidence of replacement of an aminooxy group by a cyclic secondary amide, suggesting that the mechanism of the unexpected resin cleavage is likely more complex than direct transamidation. This unusual and potentially novel reaction warrants further study, especially in the context of the growing use of aminooxy acids as peptidomimetics.

Whatever the mechanism of the unexpected cleavage, the reaction was fortuitous in that C-terminal cyclic tertiary amides would have been low on the list of new hairpin polyamide tails to synthesize. Firstly, there is no longer a proton on the amide for hydrogen bonding with the minor groove acceptor functionalities. Secondly, several studies have demonstrated that steric bulk at the C-terminal side of a hairpin polyamide is deleterious for binding and specificity.<sup>32-35</sup> However, in retrospect, all of these studies place substituents on the N-terminus, which presumably clash with the  $\beta$ -Dp tail. The loss of a hydrogen bond may be compensated by favorable Van der Waals interactions within the minor groove. The difference between morpholine and piperidine may be a negative steric interaction of the extra set of piperidine hydrogens, or the result of conformational preferences of the ring as a whole. As expected, piperazine is preferred

by the negatively charged minor groove. The toleration of cyclic non-aromatic six-membered rings at the C-terminus raises allows one to envision hairpin dimers and conjugates with significantly different kinds of linkages than were previously envisioned. It is expected that the C-terminal cyclic tertiary amide polyamides will be extremely selective for A,T against G,C at the N-1 position (see Figure 5.1). On the other hand, they may be compact enough not to interact with the N-2 position. This would be a very desirable set of properties, and warrants further investigation. It will also be interesting to see if C-terminal acyclic tertiary amides, such as dimethyl- or diethylamide would bind with similar affinity. Also of note is that given their binding site on the HER2 promoter, compounds **16-18** are expected to inhibit both TBP- (Chapter 3) and ESX-binding.<sup>22</sup> These potential inhibitors of HER2 transcription may display different uptake and nuclear localization properties than standard  $\beta$ -Dp hairpins, while maintaining favorable DNA binding properties. While the aminooxy route to the new C-terminal tails is interesting in-and-of-itself from a chemical point of view, from a practical point of view the development of polyamide synthesis on oxime resin should allow for the facile production of these and other unexplored derivatives.

## Experimental

Carboxymethoxylamine hemihydrochloride, morpholine, and piperazine were purchased from Aldrich. All other materials were as previously described.<sup>9,19,20</sup> High resolution electrospray ionization mass spectrometry was carried out at the University of California, Los Angeles Mass Spectrometry Facility. All other procedures and equipment were as described in Chapter 5A.

**Boc-Pyrrole-Aminooxy acid (13).** 1,2,3-Benzotriazol-1-yl 4-[(*tert*-butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate<sup>9</sup> (**5**) (3.56 g, 10 mmol) and carboxymethoxylamine hemihydrochloride (1.09 g, 10 mmol) were dissolved in DMF (50 mL).



DIEA (7 mL, 40 mmol) was added and the mixture was stirred at room temperature for 10 hours. The solution was concentrated *in vacuo* and the resulting oil was partitioned between 250 mL saturated NaHCO<sub>3</sub> and 250 mL ether. The aqueous layer was acidified to pH 3 with concentrated HCl, and extracted with EtOAc (2 x 600 mL). The combined ethyl acetate extracts were concentrated *in vacuo* to 200 mL and washed with 0.1 N HCl (2 x 400 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, and concentrated *in vacuo* to yield **4** as a yellow oil (2.64 g, 8.43 mmol, 84% yield): <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>) δ 11.40 (s, 1H), 9.06 (s, 1H), 6.93 (s, 1H), 6.57 (s, 1H), 4.40 (s, 2H), 3.71 (s, 3H), 1.41 (s, 9H); EIMS *m/e* 313.1278 (313.1274 calc. for C<sub>13</sub>H<sub>19</sub>N<sub>3</sub>O<sub>6</sub>).

**ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONH<sub>2</sub> (5).** Boc-β-Ala-PAM resin (1 g, 0.25 mmol/g) was deprotected with 80% TFA/DCM and 0.4 M PhSH. Boc-pyrrole-hydroxamic ester **4** (313 mg, 1 mmol) was activated by treatment with DCC (206 mg, 1 mmol) and *N*-hydroxysuccinimide (116 mg, 1 mmol) in 2 mL DMF for one hour. The solution was filtered and added to the resin, 1 mL of DIEA was added and the mixture was shaken at room temperature for six hours to yield Boc-Py-CONHOCH<sub>2</sub>CO-β-Ala-PAM resin (**14**), which was elaborated to ImImPy-(R)<sup>FmocHN</sup>γ-ImPyPyCONHOCH<sub>2</sub>CO-β-Ala-PAM resin in a stepwise fashion by manual solid phase synthesis.<sup>9,19</sup> A sample of the polyamide resin (155 mg) was treated with 160 mg Pd(OAc)<sub>2</sub> in 2 mL DMF. The mixture was shaken at 37° C for 2 hours, 600 mg of HCO<sub>2</sub>NH<sub>4</sub> in 900 μL H<sub>2</sub>O was added, and shaking at 37° C continued for seven hours. The reaction mixture was filtered and diluted to 8 mL with 0.1% (w/v) TFA, and the crude polyamide was purified by reverse-phase HPLC as previously described.<sup>9</sup> ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPyCONH<sub>2</sub> (302 μg, 360 nmol, 1.2% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder. Spectroscopic analysis confirmed this product was identical to **5** produced by the oxime resin route (Chapter 5A).

**ImPyPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPyPyCO-piperidine (16).** Resin **14** was elaborated to ImPyPyPy-(R)<sup>FmocHN</sup>γ-PyPyPyPyCONHOCH<sub>2</sub>CO-β-Ala-PAM resin (**15**) in a stepwise fashion by manual solid phase synthesis.<sup>9,19</sup> A sample of polyamide resin **15** (100 mg) was treated with 1 mL piperidine in 2 mL DMF. The mixture was shaken at room temperature for one hour. The reaction mixture was filtered and concentrated *in vacuo*, then purified by reverse-phase HPLC as previously described.<sup>9</sup> ImPyPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPyPyCO-piperidine (92 μg, 80 nmol, 0.42% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder: UV (H<sub>2</sub>O) λ<sub>max</sub> 312 (69520); MALDI-TOF-MS 1148.60 (1148.52 calc for [M+H] C<sub>56</sub>H<sub>65</sub>N<sub>19</sub>O<sub>9</sub>).

**ImPyPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPyPyCO-morpholine (17).** A sample of polyamide resin **15** (100 mg) was treated with 1 mL morpholine in 2 mL DMF. The mixture was shaken at room temperature for one hour. The reaction mixture was filtered and concentrated *in vacuo*, then purified by reverse-phase HPLC as previously described.<sup>9</sup> ImPyPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPyPyCO-morpholine (121 μg, 105 nmol, 0.55% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder: UV (H<sub>2</sub>O) λ<sub>max</sub> 316 (69520); MALDI-TOF-MS 1150.60 (1150.50 calc for [M+H] C<sub>55</sub>H<sub>63</sub>N<sub>19</sub>O<sub>10</sub>).

**ImPyPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPyPyCO-piperazine (18).** A sample of polyamide resin **15** (100 mg) was treated with 3 mL 1 M piperazine in DMF. The mixture was shaken at room temperature for one hour. The reaction mixture was filtered and concentrated *in vacuo*, then purified by reverse-phase HPLC as previously described.<sup>9</sup> ImPyPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPyPyCO-piperazine (138 μg, 120 nmol, 0.63% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder: UV (H<sub>2</sub>O) λ<sub>max</sub> 314 (69520); MALDI-TOF-MS 1149.50 (1149.52 calc for [M+H] C<sub>55</sub>H<sub>64</sub>N<sub>20</sub>O<sub>9</sub>).

**Quantitative DNase I Footprinting.** As previously reported,<sup>23</sup> for DNA fragments derived from the HER2 promoter<sup>22</sup> and plasmid pJT8.<sup>24</sup>

## Acknowledgements

We are grateful to the NIH for research support and to the Ralph M. Parsons Foundation for a pre-doctoral fellowship to J.M.B. We thank Tom Minehan, Anna Mapp, and Paul Foreancig for helpful advice and discussions concerning the synthetic aspects of this work. We are especially grateful to Doan Nguyen and Nick Wurtz, whose interest in the project led to the further developments described in Chapter 5A.

## References and Notes

- (1) Compound numbering follows from Chapter 5A.
- (2) Yang, D.; Li, B.; Ng, F.-F.; Yan, Y.-L.; Qu, J.; Wu, Y.-D. *J. Org. Chem.* **2001**, *66*, 7303-7312.
- (3) Peter, C.; Daura, X.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2000**, *122*, 7461-7466.
- (4) Shin, I.; Lee, M.; Lee, J.; Jung, M.; Lee, W.; Yoon, J. *J. Org. Chem.* **2000**, *65*, 7667-7675.
- (5) Yang, D.; Qu, J.; Li, B.; Ng, F.-F.; Wang, X.-C.; Cheung, K.-K.; Wang, D.-P.; Wu, Y.-D. *J. Am. Chem. Soc.* **1999**, *121*, 589-590.
- (6) Yang, D.; Ng, F.-F.; Li, Z.-J.; Wu, Y. D.; Chan, K. W. K.; Wang, D.-P. *J. Am. Chem. Soc.* **1996**, *118*, 9794-9795.
- (7) (a) Modeling performed by Dr. Paul Floreancig based on the NMR structure described in (b).  
(b) deClairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1999**, *121*, 2956-2964.

- (8) Polyamide **12** did appear to bind to G at the aminooxy acid position in some cases, as well as A and T. However, the specificity of the other residues in the polyamide was reduced. In particular, the  $\gamma$ -turn and pyrrole rings adjacent to the  $\gamma$ -turn showed an increased affinity for G,C base pairs. Polyamide **12** coated the DNA (bound non-specifically to all sequences) at high concentrations, but at lower concentrations it did show specificity, which in most cases could be rationalized by the pairing rules with a decrease in specificity at the tail side of the polyamide. In other cases, polyamide **12** was likely responding to the microstructure of the DNA in a way that was not readily obvious from the sequence of either the DNA or the polyamide.
- (9) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
- (10) Keck, G. E.; McHardy, S. F.; Wager, T. T. *Tetrahedron Lett.* **1995**, *36*, 7419-7422.
- (11) Canne, L. E.; Bark, S. J.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118*, 5891-5896.
- (12) Meloni, M.; Taddei, M. *Organic Letters* **2001**, *3*, 337-340.
- (13) Myers, R. M.; Langston, S. P.; Conway, S. P.; Abell, C. *Org. Lett.* **2000**, *2*, 1349-1352.
- (14) Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Org. Lett.* **2001**, *3*, 1201-1203.
- (15) Wurtz, N. R. Ph.D. Dissertation, California Institute of Technology: Pasadena, CA, 2002.
- (16) Kurth, M.; Pelegrin, A.; Rose, K.; Offord, R. E.; Pochon, S.; Mach, J.-P.; Buchegger, F. *J. Med. Chem.* **1993**, *36*, 1255-1261.
- (17) Bark, S. J.; Schmid, S.; Hahn, K. M. *J. Am. Chem. Soc.* **2000**, *122*, 3567-3573.
- (18) Canne, L. E.; Ferre-D'Amare, A. R.; Burley, S. K.; Kent, S. B. H. *J. Am. Chem. Soc.* **1995**, *117*, 2998-3007.

- (19) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.
- (20) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121-1129.
- (21) Carpino, L. A. *Acc. Chem. Res.* **1987**, *20*, 401-407.
- (22) Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A. *J. Biol. Chem.* **2000**, *275*, 24246-24254.
- (23) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.
- (24) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561.
- (25) Smith, M. E.; Adkins, H. *J. Am. Chem. Soc.* **1938**, *60*, 657.
- (26) Stirling, C. J. M. *J. Chem. Soc.* **1958**, 5431.
- (27) Bon, E.; Bigg, D. C. H.; Bertrand, G. *J. Org. Chem.* **1994**, *59*, 4035-4036.
- (28) Jencks, W. P.; Gilchrist, M. *J. Am. Chem. Soc.* **1964**, *86*, 5616-5620.
- (29) Howard, J. C.; McPherson, J. C.; Chuang, A. H.-L. *J. Med. Chem.* **1974**, *17*, 236-238.
- (30) Yang, D.; Kim, S.-H.; Kahne, D. *J. Am. Chem. Soc.* **1991**, *113*, 4715.
- (31) Canne, L. E.; Ferre-D'Amare, A. R.; Burley, S. K.; Kent, S. B. H. *J. Am. Chem. Soc.* **1995**, *117*, 2998-3007.
- (32) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147-6152.
- (33) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756-8765.
- (34) Hawkins, C. A.; de Clairic, R. P.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **2000**, *122*, 5235-5243.
- (35) Minehan, T. G.; Dervan, P. B. Unpublished results.