DNA-Templated Dimerizations of Minor Groove–Binding Polyamides

Thesis by

Adam Thomas Poulin-Kerstien

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2005

(Defended 17 May 2005)

…for Katherine, my love…

Acknowledgments

I would first and foremost like to thank my family for all their support and encouragement over the past five years. Katherine, you've been an emotional and scientific inspiration. You continue to enrich my life in countless ways. To my parents, Tom and Jane, and my sister and brother, Lissa and Darren, it has been wonderful interacting with you all as adults and friends, and no longer just as family members. I feel very fortunate to have spent these several years in close proximity. I value all the times we've shared together as a result. To my parents-in-law, John and Anne, and my sisters-in-law, Jess and Claire, thank you for your support and encouragement and all the good times we've had…torn ACL aside.

I would like to thank my research advisor, Peter Dervan, for providing me with the means and opportunity to pursue exciting research directions. Your guidance often proved invaluable. I feel very fortunate to have worked for you. Your insights, outlook, and mentoring style will continue to influence me for years to come. I would also like to thank the members of my committee, Professors Steve Mayo, Bob Grubbs, and Linda Hsieh-Wilson for their support and guidance.

I would like to thank the members of the Dervan lab, past and present, for providing an ideal working environment. I would particularly like to thank Doan Nguyen and Phillip Weyermann who acted as my day-to-day scientific mentors. I could not have accomplished what I did without their help. Tim Best, Nick Nickols, Dave Chenoweth, and Mike Brochu were my lab mates during my time here, and I could not have asked for better people to share space with. Thanks for making lab an encouraging and interesting place to be. I would also particularly like to thank the other members of the Dervan lab

who helped make the last five years enjoyable and productive: Jason Belitsky, Nick Wurtz, Christophe Briehn, Adam Urbach, Victor Rucker, Shane Foister, Clay Wang, Jim Sanchez, Ryan Stafford, Justin Cohen, Sherry Tsai, Carey Hsu, Jim Puckett, Mike Marques, Ray Doss, Steve Fiacco, Claire Jacobs, Michelle Farkas, Eric Fechter, Julie Poposki, and those already mentioned. I also want to thank all the people with whom I've been fortunate enough to share time outside lab. Your friendships are what make these 5 years worth remembering.

I would also like to thank my early mentors, whose influence eventually led me to chemistry. These include my high school chemistry teacher, Dr. Pinkerton, Professor Mark Marshall at Amherst College, and my undergraduate advisor, Professor David Hansen. You all possess the rare gift for conveying information in an engaging way, and you are all an inspiration.

Finally, I would like to dedicate the work in this thesis in part to Ben Edelson, whose memory lives on in each of those he touched.

Abstract

Polyamides have emerged as a class of small molecules capable of binding the minor groove of DNA with high affinity and sequence specificity that have potential applications in molecular biology and human medicine. In efforts towards the use of polyamides in living cells, we report research directed towards DNA-templated formations of polyamide dimers. We find that formation of polyamide dimers, linked both turn-to-turn and turn-to-tail, can be templated via a 1,3-dipolar cycloaddition using a targeted sequence of DNA. The dimer products formed *in situ* may prove to have interesting biological effects.

Also reported in this thesis are several uses of polyamides as molecular tools. We find that polyamide-biotin conjugates are able to selectively bind and capture targeted pieces of DNA via streptavidin-coated magnetic beads, effectively enriching mixtures of DNA fragments in the fragment of interest. Such molecules may find utility in the identification of DNA-protein complexes. In a second utility we report the use of polyamide-maleimide and chlorambucil conjugates to impart sequence specificity on nonspecific DNA enzymes for crystallographic studies.

Table of Contents

List of Figures and Tables

Chapter 3

Chapter 4

Appendix I

Chapter 1

Introduction: The Evolution of DNA-Binding Small Molecules

Abstract

During the past 20 years, polyamides have evolved from the natural product distamycin to a new class of programmable heterocyclic oligomers that bind a broad repertoire of DNA sequences with high affinity and specificity. This chapter details recent advances in this field of research, focusing on molecular recognition of DNA, and biological applications such as modulating gene expression by small molecules.

Background and Significance.

Sequence-Specific Recognition of DNA: From Natural Roots.

The natural product distamycin contains three *N*-methylpyrrole (Py) amino acids and binds in the minor groove of DNA at A,T tracts 4–5 base pairs (bp) in size.[1, 2] Distamycin inhibits DNA dependent processes, including transcription, and has antibacterial,[3] antimalarial,[4] antifungal,[5] and antiviral activities,[6] but is of limited use because of toxicity.[7]
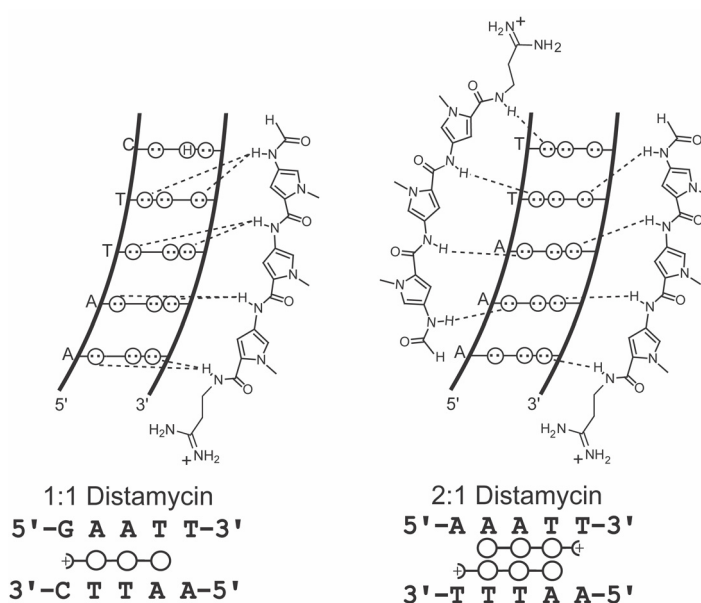
Figure 1.1. Schematic representation of the two DNA-binding modes of distamycin. Hydrogen bonds are shown as dashed lines. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Open circles below represent pyrrole amino acid units.

Structural studies of distamycin•DNA complexes revealed that the crescent-shaped molecule binds A,T tracts in both 2:1 and 1:1 ligand:DNA stoichiometries (Figure 1.1).[8-10] These structures revealed key ligand-DNA interactions, such as a series of hydrogen bonds between pyrrole carboxamides and the edges of the nucleobases on the adjacent DNA strand. These studies revealed that both the shape complementarity and the specific hydrogen-bonding profile of distamycin account for its affinity and

specificity toward B-form DNA. Over the nearly 20 years since the original structural work, analogues have been created and characterized that bind a large number of different DNA sequences in a predictable fashion.[11] We now have a set of 5- and 6- membered heterocyclic amino acids that can be combined as modular, antiparallel ring pairs in the minor groove of DNA to recognize predetermined sequences of DNA with affinities and specificities comparable to DNA-binding proteins.[12] Presented here are recent advances in the field of DNA recognition by polyamides.

The Pairing Rules.

In a formal sense, the four Watson-Crick base pairs can be differentiated on the minor groove floor by the specific positions of hydrogen bond donors and acceptors, as well as by differences in molecular shape and electronic potential surfaces (Figure 1.2a).[12] The exocylic amine of guanine presents an unsymmetrical hydrogen bond donor "bump" on the minor groove edge of a G•C base pair. A key study in the early 1990s demonstrated that the *N*-methylimidazole (Im)-containing polyamide ImPyPy bound to the five bp sequence 5'-WGWCW-3' (where W = A or T).[13] This result was rationalized in terms of the formation of a 2:1 polyamide-DNA complex in which an antiparallel ring pairing of Im stacked against Py could specifically distinguish a G•C from the other three base pairs (Figures 1.2 and 1.3).

The Im/Py pair has been explored by extensive studies, including analyses of binding in hundreds of different sequence contexts. Crystal structures confirmed the existence of a hydrogen bond between the Im nitrogen and the exocyclic NH_2 of guanine when the Im/Py pair binds opposite the G•C base pair.[14] The preference for a linear
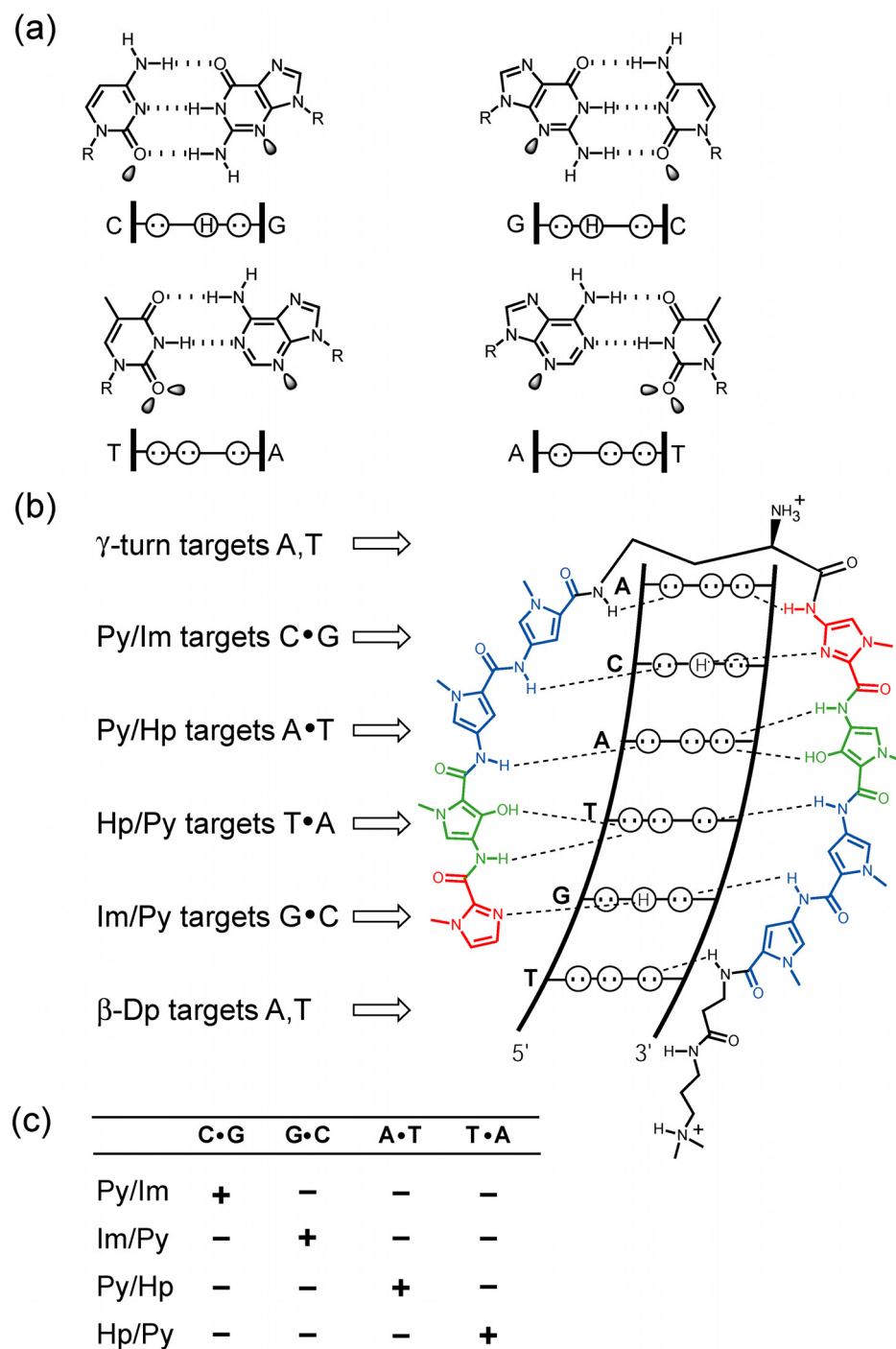
Figure 1.2. (a) Structures of the four Watson-Crick base pairs. The R group represents the sugar-phosphate backbone of DNA, and shaded orbitals represent electron lone pairs projecting into the minor groove. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles with an H represent the exocyclic 2-amino group of guanine. (b) Schematic model for the hairpin polyamide ImHpPyPy-γ-ImHpPyPy-β-Dp bound to its match site 5'-TGTACA-3' as determined by the pairing rules for recognition of all four Watson-Crick base pairs of DNA in the minor groove by polyamides. Putative hydrogen bonds are indicated as dashed lines. (c) Table indicating the code for minor groove recognition by polyamides. Plus and minus signs indicate favored and disfavored interactions, respectively.

hydrogen bond, coupled with the unfavorable angle to an Im over the cytosine side of the

base pair, provides a basis for the ability of an Im/Py pair to discriminate G•C from C•G

(Figure 1.3). Thermodynamic investigations dissected binding free energies into

enthalpic and entropic contributions, revealing that the sequence selectivity of the Im/Py

pair is driven by a favorable enthalpic contribution.[15]
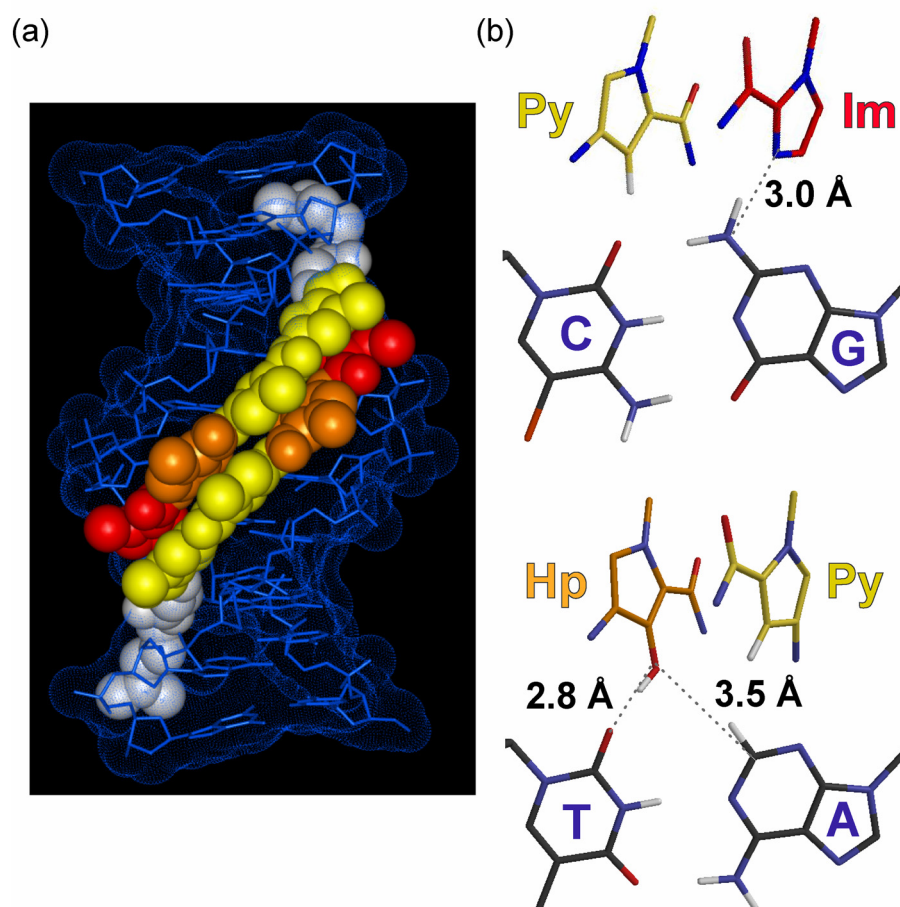
Figure 1.3. (a) X-ray crystal structure of ImHpPyPy-*β*-Dp (Dp = dimethylaminopropylamine) bound in a 2:1 complex with its target DNA site, 5'-AGTACT-3' (PDB code 407D). Im residues are red, Hp residues are orange, Py residues are yellow. (b) Detail of the Py/Im pair interacting with the C•G base pair (top) and of the Hp/Py pair interacting with the T•A base pair. Dashed lines indicate interatomic distances between carbon, nitrogen, and oxygen atoms emphasizing the close interactions responsible for specificity.

Discrimination of the A, T base pairs was not achieved until relatively recently.

The A•T base pair appears fairly symmetrical, with both adenine and thymine presenting

a hydrogen bond acceptor to the floor of the minor groove (Figure 1.2a). However, closer inspection reveals that a small asymmetric cleft is formed between the thymine O2 and adenine C2. Furthermore, the N3 of adenine presents only one lone pair while the O2 of thymine presents two lone pairs capable of hydrogen bonding. Informed by high resolution crystallographic data from a polyamide-DNA complex, *N*-methyl-3-hydroxypyrrole (Hp) was designed, and subsequently proved to be a thymine-selective recognition element when paired across from Py (Figures 1.2 and 1.3).[16]

Crystal structures of two different Hp-containing polyamides, as their 2:1 complexes with DNA, have been determined at high resolution.[17, 18] The specificity of an Hp/Py pair was shown to arise from a combination of specific hydrogen bonds between the hydroxyl and the thymine O2, along with shape-selective recognition of the asymmetric cleft (Figure 1.3). Hp polyamides bind with lower affinities than their Py counterparts,[16] and a recent computational study argues that desolvation of the hydroxyl group upon insertion into the minor groove accounts for the energetic penalty.[19] Together, three rings–Py, Im, and Hp–can be combined as unsymmetrical pairs to recognize specifically each of the four Watson-Crick base pairs; Im/Py is specific for G•C and Hp/Py for T•A (Figure 1.2b). These interactions can be conveniently described as pairing rules (Figure 1.2c). The pairing rules should be considered as guidelines only. Antiparallel polyamide dimers bind B-form DNA, and there are limitations regarding sequences targeted due to the sequence-dependent microstructure of DNA.

Expanding and Refining DNA Recognition.

Improving Affinity and Specificity.

Covalent linkage of the two antiparallel polyamide strands results in molecules with increased affinity and specificity (Figure 1.4a). Currently, the "standard" motif is the 8-ring hairpin, in which a *γ*-aminobutyric acid linker (*γ*-turn) connects the carboxylic terminus of one polyamide to the amino terminus of its antiparallel partner.
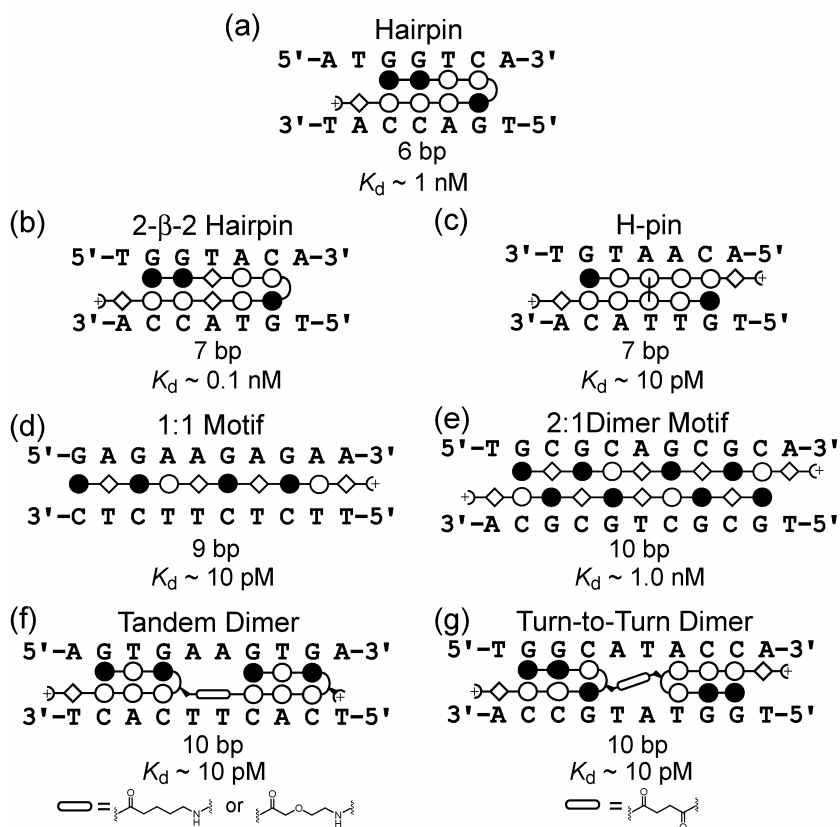
Figure 1.4. Polyamide-DNA binding motifs with approximate dissociation constants (K_d). (a) Hairpin: standard motif targeting 6 bp with high affinity and sequence specificity. (b) *β*-Ala-containing polyamides are able to bind longer sequences because the flexible aliphatic residue relaxes the curvature of the polyamide. (c) 2:1 Complex: *β*-containing polyamides are able to target long stretches of DNA as either 2:1 complexes (shown here) or 1:1 complexes depending on stoichiometry. (d) H-pin: covalent linkage of two polyamide strands is achieved at a position that is sequence-neutral. (f, g) Hairpin Dimers: linked either turn-to-tail (f) or turn-to-turn (g), these molecules are able to target long sequences of DNA with moderate specificity and high affinity. Optimized linkers for each motif are shown below. The black and open circles represent Im and Py rings, respectively; diamonds represent *β*-alanine residues; and plus signs next to diamonds represent Dp residues. A curved line connecting the sides of two circles represents the *γ*-aminobutyric acid turn, and a curved line with a wedge and a plus sign represents the chiral (*R*)$^{\text{H2N}}$*γ*-turn. For the H-pin, curved lines connecting the centers of two circles represent alkyl linkers attached to the N-methyl positions of the aromatic rings.

Compared to the unlinked homodimers, hairpin polyamides display ~100-fold higher affinity, with the γ-turn demonstrating selectivity for A,T over G,C base pairs (Figure 1.2b), presumably due to a steric clash between the aliphatic turn unit and the exocyclic amine of guanine.[20] 8-ring hairpins, which bind six bp, were shown to have affinities and specificities similar to DNA-binding proteins (i.e., K_d < 1 nM).[21] NMR studies confirmed that the γ-turn locks the register of the ring pairings, preventing the ambiguity of slipped dimers.[22] Hairpin polyamides retain the orientation preferences of unlinked antiparallel polyamides, aligning N→C with respect to the 5'→3' direction of the adjacent DNA strand.[23] The β-alanine-dimethylaminopropyl amine (β-Ala, Dp, respectively) tail substituent at the C-terminus of many hairpin polyamides is an A,T specific element (Figure 1.2b), again presumably due to a negative interaction between the aliphatic chain and the exocyclic amine of guanine.[24] The tail is thought to play a role in the orientation preference of polyamides.[25]

For some hairpins, however, "reversed binding" (a C→N alignment of the polyamide with respect to the 5'→3' direction of the adjacent DNA strand) has been observed as the preferred orientation.[23] By introducing an amino substituent at the α position of the γ-turn, reversed binding is disfavored because of a steric clash between the amino substituent and the floor of the minor groove.[26] Not only does the chiral turn maintain the specificity of hairpins, it increases the overall affinity, either by the addition of a positive charge, which interacts favorably with the negatively charged backbone or minor groove floor of the DNA polymer, or by sterically reducing the conformational freedom of the polyamide in solution.[26]

In other cases, polyamides containing aliphatic residues such as the γ-turn or β-

Ala have been shown to favor an extended 1:1 binding mode (Figure 1.4d). Depending on the stoichiometry, the *same* polyamide may bind *different* sequences.[27] For example, a hairpin containing internal *β*-alanines might bind in two different confirmations, a "folded" versus an "extended" conformation. An amino substituent on the γ-turn also disfavors the extended binding mode, and serves to lock the polyamide into the hairpin conformation.[28] A larger, N-acetyl group increases this effect, with an 8-ring, N-acetyl-substituted polyamide favoring the hairpin conformation over extended binding by >25,000-fold. Substitution of *β*-Ala for Py has also been shown to influence the tendency of a polyamide to bind in a hairpin conformation.[29] In each case, it is presumably a negative interaction between the substituent and the wall of the minor groove that drives the equilibrium toward the hairpin conformation.

Binding-Site Size.

While the standard 8-ring hairpin recognizes DNA with high affinity and specificity, it targets only six bp. For biological applications, binding site size may be critical because longer sequences would be expected to occur less frequently in the genome. Early attempts to increase the targeted site size by simply extending the number of aromatic rings resulted in polyamides with decreased affinity.[30] Crystal structures of polyamide/DNA complexes have shown that the polyamide rise-per-residue matches the pitch of the B-DNA helix–that is, the spacing of the pyrrole and imidazole rings matches the spacing of the DNA base pairs.[14, 17, 18] However, the inherent crescent-shaped curvature of polyamides is slightly tighter than the curvature of the minor groove. Beyond five contiguous rings, the shape of a polyamide is no longer complementary to

DNA, and the resulting loss of specific contacts accounts for the observed loss in affinity and specificity.[14]

The hypercurvature of polyamides can be overcome by the introduction of the flexible β-alanine residue. When introduced as a Py replacement,[31] its flexibility relaxes the curvature of the polyamide, restoring complementarity to DNA. β-Ala-containing polyamides have been used to target up to 9 bp with high affinity and specificity (Figure 4b, d, e).[32]

Another strategy to target longer sequences of DNA is to covalently link two hairpin polyamides with a flexible linker. These dimeric polyamides, linked both turn-to-tail and turn-to-turn, have been shown to bind longer sequences with high affinity (Figure 1.4f, g).[33-35] Both turn-to-turn and turn-to-tail dimers with optimized linkers showed good selectivity for a 10 bp site (over 11 and 12 bp sites) but exhibited poor specificity (expressed in terms of affinity for match over single base pair mismatch sites). Nonetheless, an impressive application of the tandem motif was demonstrated by Laemmli and coworkers, who used tandem-hairpins to stain insect or vertebrate telomeres, $(TTAGG)_n$ or $(TTAGGG)_n$ repeats, respectively, with remarkable selectivity in fixed cells and chromosome spreads.[36]

H-Pin and U-Pin Motifs.

Polyamides also can be linked, via the ring nitrogens, with an alkyl spacer that projects away from the minor groove. When placed in the center of a polyamide, the resultant branched molecule has been termed an H-pin (Figure 1.4c); when placed at the end, a U-pin. H-pin polyamides bind with high affinity and good specificity.[37] Recent

efforts to improve the synthetic methods for H-pins using alkene metathesis on a solid support have enabled a detailed study of the optimal alkyl linker length, demonstrating that four and six methylene units provide the highest affinities.[38] U-pin polyamides behave similarly.[39] The affinity of an 8-ring U-pin is most comparable to a 6-ring hairpin polyamide, likely due to a loss of two hydrogen bond donors upon removal of the γ-turn element. Thus, the dimeric Py-Im U-turn element may be thought of as a C•G specific replacement for the γ-turn. In combination with removal of the β-Ala-Dp tail, H-pin and U-pin polyamides could potentially bind purely G,C sites, a sequence type that has been difficult to target with other polyamide motifs.

1:1 Polyamide:DNA Complexes.

Homopurine tracts have been a challenging target for polyamide binding because these DNA sequences have particularly narrow minor grooves.[40] While the width of the minor groove of these DNA sequences may be too narrow for hairpin polyamides, a single Py-Im-β-containing strand may be accommodated at such sequences.

In 1:1 polyamide•DNA complexes, β-linked polyamides appear to prefer a single orientation, N→C with respect to the 3'→5' direction of the purine-rich strand.[41] Footprinting of 1:1 complexes has shown that Im residues do not distinguish G,C from A,T whereas Py and β residues prefer A,T over G,C base pairs.[41]

A high-resolution 1:1 solution NMR structure of ImPy-β-Im-β-ImPy-β-Dp elucidated the role of β-alanine in minor groove recognition (Figure 1.5).[42] β-Ala allows both Im rings in the β-Im-β-Im subunit to adapt to the relatively large dihedral required for hydrogen bonding. Additionally, close contacts of β-alanine to the floor of the minor

groove provided a structural explanation for its observed A/T specificity.

(a)   (b)
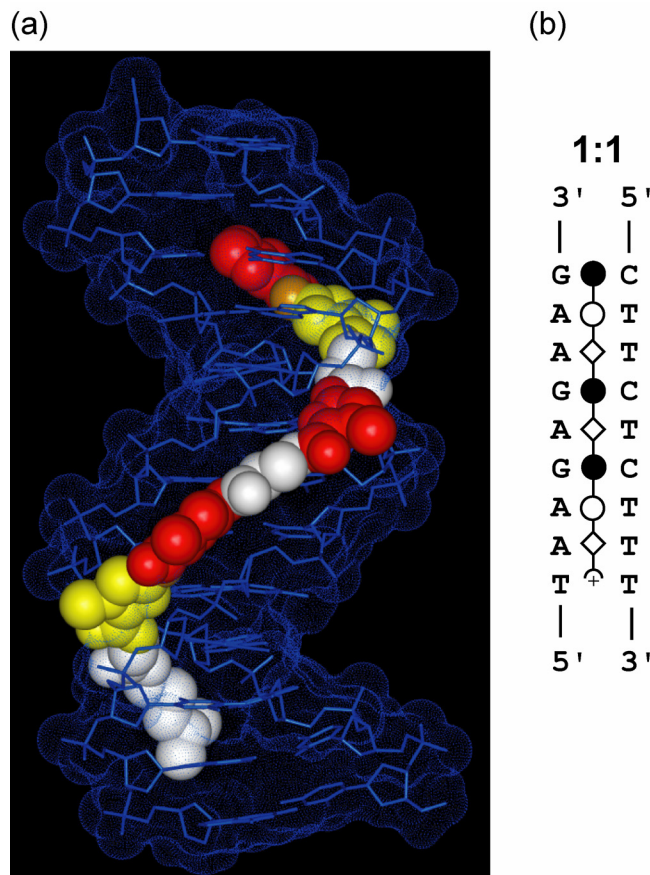
Figure 1.5. (a) NMR Structure of a 1:1 polyamide:DNA complex. The DNA backbone and bases are shown in blue. Im residues are red, Py residues are yellow, *β*-Ala and Dp residues are white. (b) Schematic representation of the structure shown in (a).

While the entire recognition code in the 1:1 motif has not been fully elucidated, this type of binding mode is uniquely suited to targeting homopurine sequences.[41, 43] Laemmli and coworkers reported a striking example of a 1:1 motif polyamide targeted to the satellite regions of *Drosophila melanogaster* being able to induce specific gain-of-function and loss-of-function phenotypes when fed to developing flies.[44]

Improving Synthetic Methodology.

The investigation of minor groove-binding polyamides was greatly accelerated by

the implementation of solid-phase synthesis.[45] Originally demonstrated on Boc-β-Ala-PAM resin with Boc-protected monomers, it was also shown that Fmoc chemistry could be employed with suitably protected monomers and Fmoc-β-Ala-Wang resin.[46] Recently, Pessi and coworkers used a sulfonamide-based safety-catch resin to prepare derivatives of hairpin polyamides.[47] Upon activation of the linker, resin-bound polyamides were readily cleaved with stoichiometric quantities of nucleophile to provide thioesters or peptide conjugates.
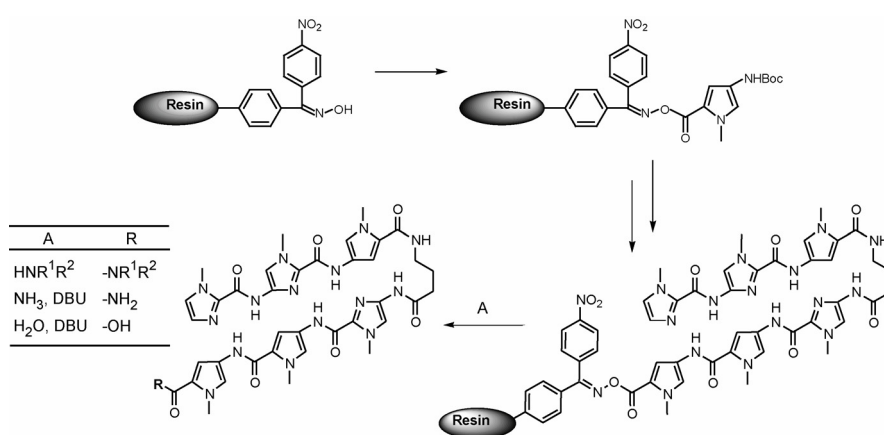
Figure 1.6. Scheme for synthesis of polyamides on the Kaiser oxime solid-support resin. Cleavage from resin with various reagents (**A**) can result in polyamides with shorter C-terminal groups than molecules prepared on β-Ala-PAM resin. The amine HNR^1R^2 may be a primary or secondary alkyl amine.

While allowing rapid preparation of a range of polyamides, these resins install a T,A selective β-Ala residue at the C-terminus, which places limits on the DNA sites that can be targeted.[24] The shortest tail available from these resins is a propanolamide, obtained by reductive cleavage. Polyamides prepared on Boc-Gly-PAM resin can be reductively cleaved to obtain ethanolamide tails, but it was expected that further truncation of the C-terminus would be necessary for tolerance of G,C at the tail position.[24] The Kaiser oxime resin was therefore adapted to polyamide synthesis, allowing the preparation of polyamides with shorter C-termini (Figure 1.6). These

molecules display the desired tolerance for G,C bases while maintaining high affinities.[48]

New Ring Systems.

The specificity of cofacial aromatic amino acid pairings is highly dependent upon their position within a given polyamide. For example, an Im/Py pairing is specific for G•C at both internal and terminal positions.[13] In contrast, Hp/Py and Py/Py pairings, while specific for T•A and A,T, respectively at internal positions, lose all specificity when incorporated at the *N*-terminal cap position. The loss of specificity at the cap position is presumably a result of conformational freedom caused by the absence of a second "groove-anchoring" carboxamide, allowing terminal rings to bind DNA in either of two rotamers.

(a)

5'-C A T [X] T A C A T A-3'
3'-G T A [Y] A T G T A T-5'

K_a values (x 10^9 M^{-1})

X·Y				
T·A	3.8	5.1	**2.0**	13
A·T	2.8	3.1	0.32	3.7
C·G	3.2	0.26	≤ 0.01	0.21
G·C	**70**	1.1	≤ 0.01	0.31

(b)
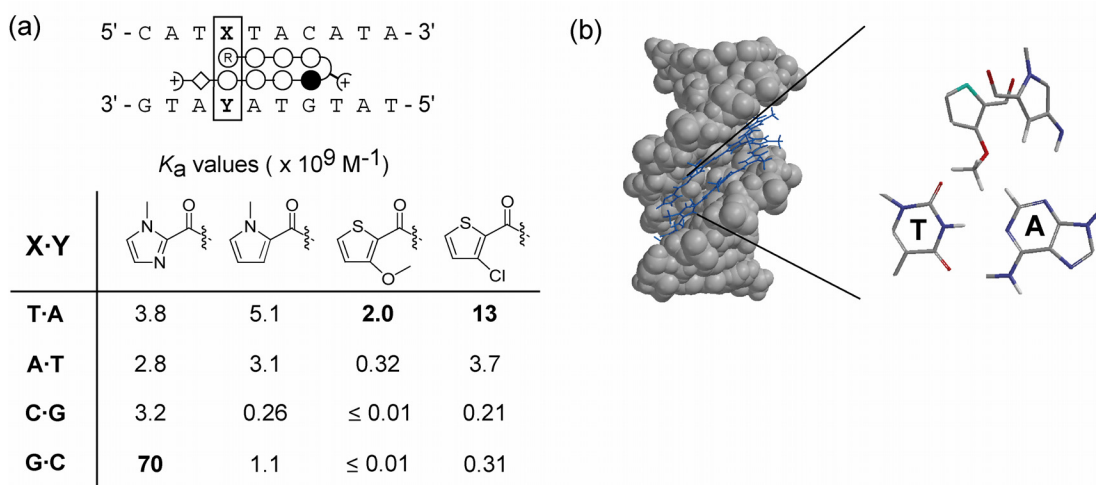
Figure 1.7. (a) Table indicating the equilibrium association constants (K_a) for polyamides containing the indicated ring at the N-terminal position of the 8-ring hairpin polyamide pictured above at sequences containing each of the four Watson-Crick base pairs. When paired with Py, Im shows a preference for guanine while 3-substituted thiophenes show a preference for thymine. (b) Model created using the PC Spartan (Wavefunction, Inc.) illustrating the shape-selective recognition of thymine by the 3-methoxy substituent of thiophene.

Recently, a library of substituted five-membered aromatic carboxylic acids was incorporated at the *N*-terminal position of an 8-ring polyamide, and *N*-terminal specificity

was probed.[49]   It was found that 3-chlorothiophene (Ct) and 3-methoxythiophene (Mt) rings, when paired against Py, imparted a moderate degree of specificity for T•A over A•T (6- and 3-fold, respectively), and a large degree of specificity over G,C base pairs (>200-fold) (Figure 1.7a).   While the Mt/Py imparts a higher degree of specificity for T•A over A•T, the Ct/Py pair imparts modest specificity as well as higher affinity. Molecular modeling of Mt- and Ct-containing polyamides indicated that the rotamer that places the 3-substituent into the floor of the minor groove is energetically favored. This substituent fills the asymmetric cleft created by the N2 of thymine and the C2 of adenine, accounting for the observed preference for T•A (Figure 1.7b).

While most polyamide research has focused on 5-membered aromatic ring systems, other scaffolds have been shown to bind DNA. The benzimidazole ring system represents a different structural framework that is amenable to functionalization on the 6-membered ring and appears to impart a curvature that is complementary to DNA.[50] Indeed, the classic minor groove-binding Hoechst dyes are composed of benzimidazole units, and a number of derivatives of these molecules have been prepared. We have incorporated benzimidazole derivatives into the backbones of hairpin polyamides in a manner that preserves critical hydrogen bonding contacts and overall molecular shape (Figure 1.8).[51, 52]   The imidazopyridine (Ip) and hydroxybenzimidazole (Hz) rings are introduced into polyamides as the dimeric subunits PyIp and PyHz, respectively, in which the Py ring is directly connected to the benzimidazole derivative without an intervening amide bond.[51, 52]   DNase I footprinting indicates that the Ip/Py and Hz/Py pairs are functionally identical, at least in some sequence contexts, to the analogous five-
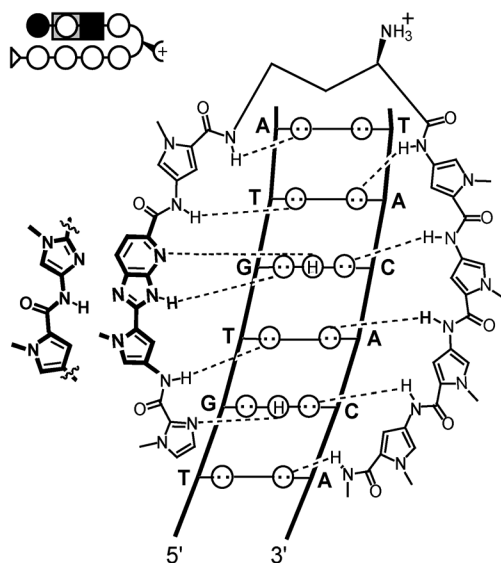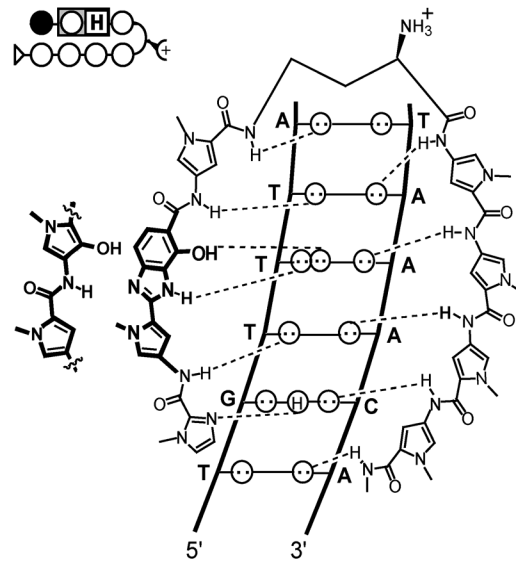
Figure 1.8. Recognition of the DNA minor groove with benzimidazole-derivatives. (a) Structure of polyamide containing Py-hydroxybenzimidazole (PyHz). (b) Structure of polyamide containing Py-imidazopyridine (PyIp). The 5-membered ring analogues of the dimeric benzimidazole derivatives are shown to the right of each model. Putative hydrogen bonds are indicated with dashed lines. Rectangles containing a white circle and a black square indicate the PyIp unit. Rectangles containing a white circle and a square with an H indicated the PyHz unit. Other symbols are defined in figures 1.2 and 1.4.

membered ring pairs Im/Py and Hp/Py.[52] Significantly, we have found that the Hp-containing polyamides can degrade over time in the presence of acid or free radicals, whereas the analogous Hz-containing compounds are chemically robust. Thus, the Hz/Py pair is a strong candidate for replacing Hp/Py in biological studies.

Secondary Effects of Polyamides.

Displacement of DNA-Binding Proteins.

Polyamides bind with high affinity to a wide range of DNA sites and can competitively displace many proteins from DNA. This can have an effect on gene expression, as DNA-binding proteins are often involved in the regulation of transcription.
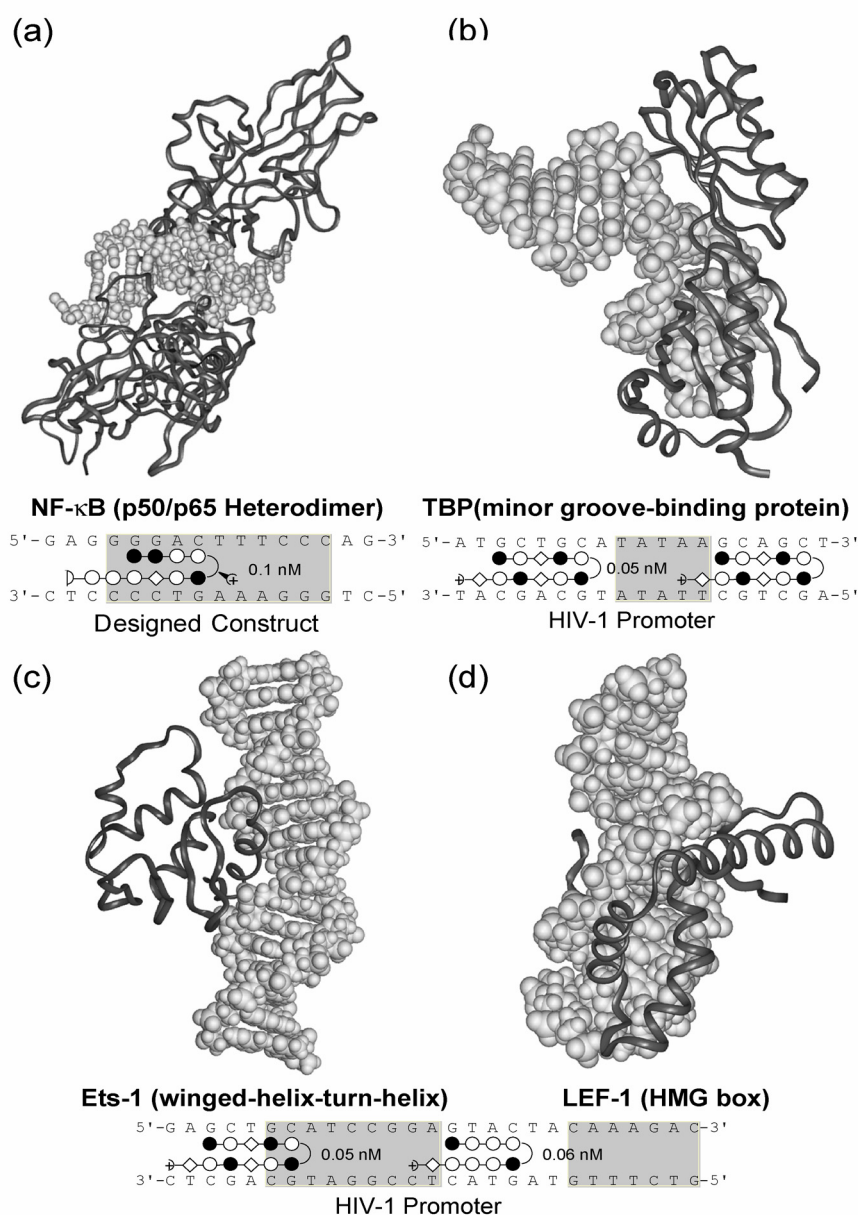
Figure 1.9. X Ray crystal structures of four different protein-DNA complexes that have been inhibited by polyamides. Below each structure is illustrated the context for inhibition, with the protein binding sites shaded and the polyamides responsible for inhibition shown bound to their match sites. Listed beside each polyamide is its dissociation constant (K_d). Symbols are as defined in figure 1.4.

One approach to modifying gene expression involves inhibition of key transcription factor (TF)-DNA complexes in a designated promoter, thus interfering with recruitment of RNA polymerases.[53] Significantly, because there are considerably fewer oncogenic TFs than potentially oncogenic signaling proteins, TF inhibition represents a uniquely

promising approach to cancer treatment.[54] The transcription factor TFIIIA was chosen as a first target because it regulates a relatively small number of genes and because the contacts between the nine zinc-finger protein and the minor groove had been established. A polyamide bound in the recognition site of TFIIIA suppressed transcription of 5 S RNA genes by RNA polymerase III *in vitro* and in cultured *Xenopus* kidney cells (Figure 1.9a).[53] Further studies used polyamides in combination with recombinant derivatives of TFIIIA subunits to elucidate essential minor groove contacts for the binding of this TF.[55]

Polyamides were then used to target viral genes transcribed by RNA polymerase II. The HIV-1 enhancer/promoter contains binding sites for multiple transcription factors, including TBP, Ets-1, and LEF-1. Two hairpin polyamides designed to bind DNA sequences immediately adjacent to the binding sites for LEF-1 and Ets-1 specifically inhibited binding of each transcription factor and prevented HIV-1 transcription in a cell-free assay (Figures 1.9b, c and d).[56] In human blood lymphocytes, treatment with the two polyamides in combination inhibited viral replication by 99%, with no significant decrease in cell viability. Inhibition of viral replication is indirect evidence for specific transcription inhibition by polyamides, because other modes of action could be involved, such as modulation of T-cell activation pathways. However, RNase protection assays indicated that the two polyamides did not alter the RNA transcript levels of several cytokine and growth-factor genes, suggesting that polyamides do affect transcription directly.

This early biological result spurred a variety of biochemical studies of the interactions of polyamides with the basal transcription machinery and TF-DNA complexes. Two studies have used promoter scanning to identify sites where polyamide

binding inhibits transcription.[57, 58]   The method uses a series of DNA constructs with designed polyamide binding sites at varying distances from the transcription start site. Essential minor groove contacts were identified for a subunit of TFIIIB (possibly TBP) in a *Xenopus* tRNA promoter,[58] as well as for TFIID-TFIIA and TBP in the HIV-1 core promoter.[57]   The binding of the homodimeric basic-helix-loop-helix TF Deadpan was investigated using a variant of promoter scanning.[59]   A series of duplex oligonucleotides based on a *Drosophila* neural promoter were designed, incorporating polyamide binding sites on different sides of the Deadpan recognition sequence and in different orientations. The TF-DNA complex was inhibited only by a polyamide binding upstream of the homodimer, establishing an asymmetric binding mode for this TF.

In the human T-cell leukemia virus type 1 (HTLV-1) promoter, polyamides targeted to G,C-rich regions flanking the viral CRE sites inhibited binding of the Tax protein and Tax transactivation *in vitro* (Figure 1.10).[60]   This example illustrates several important polyamide•DNA•protein interactions.   HTLV-1 genes are regulated by the major groove-binding protein CREB (CRE Binding Protein).   CREB-mediated transcription is enhanced by the binding of the viral protein Tax, which makes contacts with CREB and to the minor groove at sites flanking the CREB binding site. Tax then recruits CREB binding protein (CBP) via a KIX domain on CBP, which then induces transcription. Researchers found that addition of two polyamides designed to target the Tax recognition elements inhibited Tax from associating to the CREB•DNA complex, and Tax-induced transcription was abolished. Interestingly, these polyamides bind only a few base pairs away from the CRE site, yet CREB is able to co-occupy the DNA, with CREB-meditated basal transcription remaining intact. Thus, polyamides are able to

interfere very specifically with some protein•DNA interactions while leaving other nearby interactions unaffected.
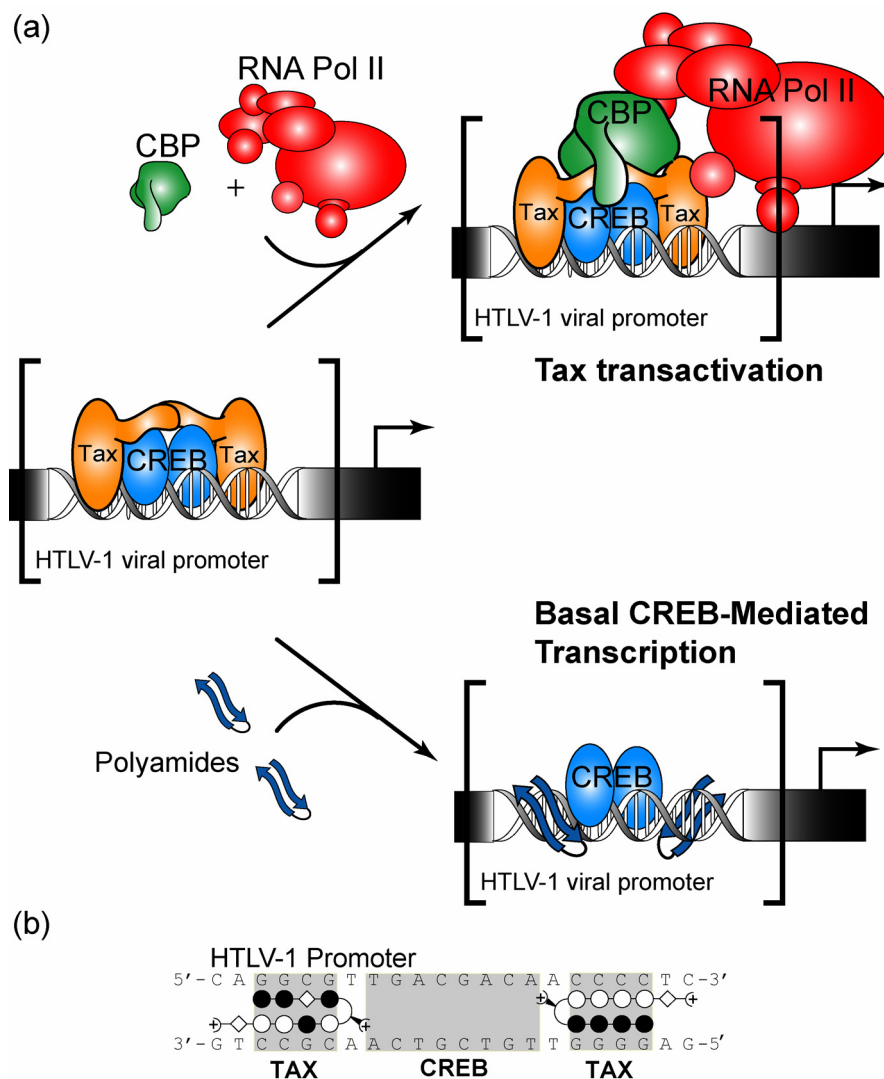
Figure 1.10. (a) Polyamide inhibition of Tax transactivation. Tax and CREB are bound to the HTLV-1 viral promoter. The trimeric complex recruits CBP and the transcriptional machinery. Polyamides specifically inhibit Tax but not CREB from binding to DNA, thus abolishing Tax transactivation while leaving basal CREB-mediated transcription unaffected. (b) Model of the sequence recognized by Tax and CREB with the structures of the polyamides used to inhibit Tax shown bound to their targeted sequences. All symbols are as defined in figure 1.4, with Tax and CREB binding sites shaded.

Several other protein•DNA interactions have been inhibited with polyamides. Bacterial gyrase recognizes a short 5'-GGCC-3' site, and a polyamide targeted to this sequence inhibited gyrase-catalyzed strand cleavage at nanomolar concentrations.[61] NF-

κB is a TF crucial for development, viral expression, inflammation, and anti-apoptotic responses. The most common form is a p50-p65 heterodimer, which binds DNA in the major groove, making several phosphate contacts throughout the binding site. Polyamides targeted to the minor groove opposite p50, but not p65, inhibit DNA binding by NF-κB (Figure 1.9a).[62] In a different study, polyamides were shown to bind very near the 3' processing end of Moloney murine leukemia virus (M-MuLV) long terminal repeat (LTR) sequences, thereby inhibiting retroviral integration catalyzed by M-MuLV Integrase (IN).[63]

The binding of Ets-1 to the HIV-1 enhancer was examined in greater detail, and polyamides were shown to inhibit the formation of a ternary Ets-1–NF-κB–DNA complex.[64] Ets-1 is a winged-helix-turn-helix TF, and its key phosphate contacts on either side of the major groove can be disrupted by a polyamide in the adjacent minor groove. The report provided evidence for cooperative DNA binding by Ets-1 and NF-κB to the HIV-1 enhancer sequence. A different Ets binding site in the HER2/*neu* promoter was targeted with hairpin polyamides that successfully blocked Ets•DNA complex formation and transcription of the HER2/*neu* oncogene in a cell-free system.[65]

Recently, researchers were able to inhibit the binding of human papilloma virus (HPV) transcription factor E2 using a tandem hairpin polyamide.[66] The E2 homodimer binds exclusively in the major groove and bends the DNA towards the body of the protein. A polyamide targeted to the E2 binding site prevents such bending, thereby destabilizing the E2•DNA complex. The topological change to the DNA caused by polyamide binding is thought to be the mechanism for E2 inhibition, and illustrates how polyamides may interfere with DNA-binding proteins without actually contacting the

protein.

Other purely major-groove binding TFs, such as the basic-region leucine zipper (bZIP) protein GCN4, have been shown to co-occupy the DNA helix in the presence of polyamides.[67]   Strategies employing polyamides functionalized with helix-distorting moieties have been successful at inhibiting such proteins. Polyamides with an attached Arg-Pro-Arg tripeptide can interfere with major-groove binding proteins by disrupting key phosphate contacts, distorting the DNA by charge neutralization, or sterically invading the major groove. An Arg-Pro-Arg polyamide conjugate successfully inhibited the binding of GCN4 to DNA,[67] and further optimization yielded a polyamide derivative with an alkyl diamine substituent that was 10-fold more potent.[68]
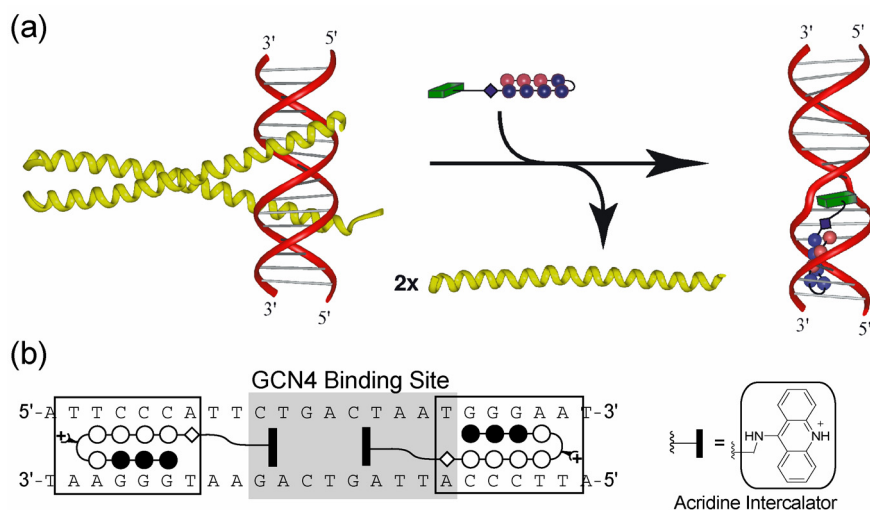
Figure 1.11. Model for allosteric inhibition of a protein-DNA complex by a polyamide-intercalator conjugate. (a) The GCN4 homodimer is displaced by the intercalating moiety of the polyamide conjugate. (b) Ball-and-stick model of the polyamide conjugate binding the target site (boxed) adjacent to the binding of the protein GCN4 (shaded). The structure of the acridine intercalator is shown at right. All other symbols are as defined in figure 1.4.

Polyamide-intercalator conjugates that distort the DNA at specific, targeted sequences by insertion of an intercalator have also proved to be highly potent inhibitors of major groove-binding proteins. Polyamides conjugated to the intercalator acridine

disrupt the DNA microstructure via unwinding, and were shown to significantly inhibit GCN4 binding when bound to sites adjacent to the GCN4 binding site, placing their acridine moieties into the GCN4 recognition element (Figure 1.11). Such molecules are promising candidates for site-selective inhibition of any DNA-binding protein.[69]

Polyamides can also upregulate transcription by inhibition of a repressor protein (derepression). For example, a hairpin polyamide was shown to block binding of the repressor IE86 to DNA, thereby upregulating transcription of the human cytomegalovirus MIEP.[70] A more complex case involves derepression of the integrated HIV-1 long terminal repeat (LTR). The human protein LSF binds in the promoter region at the LTR and recruits YY1, which then recruits histone deacetylases (HDACs). HDACs subsequently maintain LTR quiescence, which has been implicated in HIV latency, by maintaining a silent stock of pathogen. Three different live-cell models demonstrated that polyamides can inhibit LSF binding and increase expression of integrated HIV-1 promoter.[71] As with other systems, only polyamides matched to the correct protein binding site induced significant effects. Several existing drug treatments can reduce HIV-1 levels in the blood to below detectable amounts, yet the virus inevitably returns in infected patients. Derepression by inhibition of LSF-DNA binding may eventually allow HIV to be fully eradicated by drug treatments. This approach is particularly promising because LSF is a human protein, which could make the target less susceptible to resistance by HIV-1 mutations.

Recruitment of DNA-Binding Proteins

Polyamides have also been shown to affect DNA structure and function by recruiting proteins to specific, targeted sites. Most transcription factors have a DNA-binding domain and a separate domain that recruits the transcriptional machinery to that site (often called the activation domain). A polyamide can be thought of as an artificial DNA binding domain that can be linked to an activation domain. Such artificial transcription factors have been synthesized and evaluated in cell-free transcription assays.[72, 73] A hairpin polyamide tethered by a 36-atom straight-chain linker to the short (20-residue) peptide activation domain AH gives robust activation of transcription, with a size of only 4.2 kDa. Replacing the AH peptide with the shorter yet more potent activator VP2 (derived from the activator domain of the viral activator VP16) and reducing the linker from 36 atoms to eight provided a "minimal" polyamide-peptide conjugate, 3.2 kDa in size, which activated transcription slightly more effectively than the larger analogue (Figure 1.12a).[73] Since the linker length had been shown to influence activation efficiency, a set of molecules with rigid oligoproline linkers between the polyamide and the activation domain was synthesized.[74] The oligoproline linkers act as "molecule rulers," and optimal activation was observed with a Pro_{12} linker, about 36Å in length.

Many genes are influenced by multiple pathways and thus rely on the binding of several proteins. One example is the Hox (Homeobox) family of transcriptional regulators, which plays a vital role in the developmental fate of an organism. However, Hox proteins generally display poor affinity and sequence-specificity towards DNA. Instead, they are recruited to DNA by the strong, specific binding of members of the
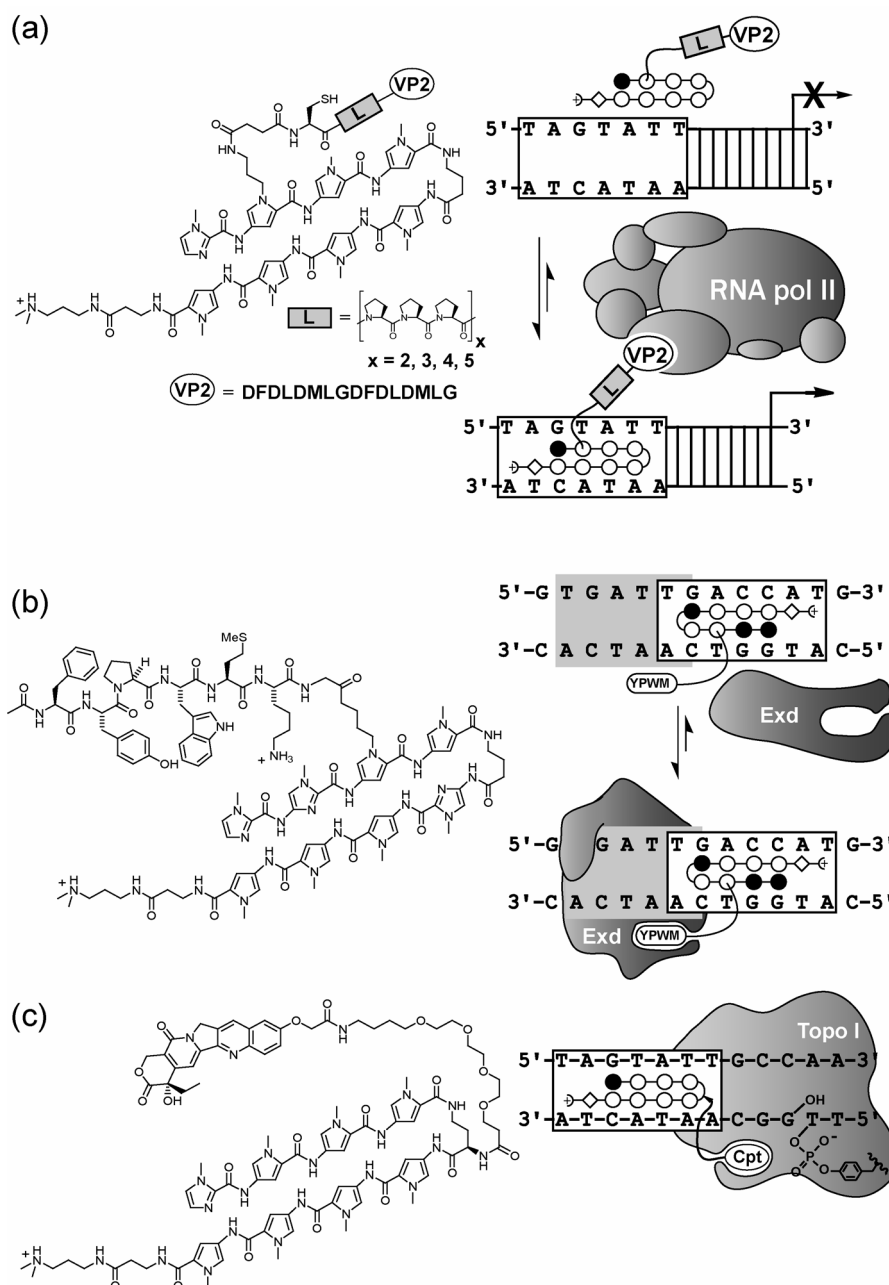
Figure 1.12. Recruitment of cellular proteins to DNA by polyamides. (a) Polyamides conjugated to the VP2 activation domain via a rigid poly-proline linker recruit the transcriptional machinery to a targeted site. (b) A polyamide functionalized with a short, YPWM peptide recruits Exd at nanomolar concentrations, changing the protein from a non-DNA-binding conformation (top) to one that binds the DNA•polyamide complex with high affinity (bottom). (c) Polyamide-camptothecin conjugates recruit topoisomerase I, inducing specific, targetable DNA strand breaks. Protein binding sites are shaded, and polyamide binding sites are boxed. All other symbols are labeled or defined in figure 1.4.

TALE (Three Amino Acid Loop Extension) class of homeodomain proteins. Recent

crystal structures of one such ternary complex shows that the Hox protein Ultrabithorax

(Ubx) interacts with the *Drosophila* TALE protein extradenticle (Exd) via a short docking YPWM peptide.[75] A polyamide functionalized with this YPWM peptide successfully recruited Exd at nanomolar concentrations, outperforming the natural Ubx protein (Figure 1.12b).[76] This demonstrates that cooperative interactions among functionalized polyamides, DNA, and a protein can stabilize the formation of a ternary complex on a composite DNA site *in vitro*.

Polyamide camptothecin conjugates specifically recruited DNA topoisomerase I (Topo I) and induced single-strand cleavage.[77] Camptothecin is known to stabilize the cleavage complex formed between a tyrosine residue on Topo I and the 3'-phosphoryl end of the DNA backbone.[78] Using polyamide-camptothecin conjugates, this cleavage complex could be generated sequence-specifically at sites adjacent to the polyamide binding site (Figure 1.12c). Since camptothecin-Topo I–DNA complexes have been shown to arrest transcription elongation,[79] polyamide-camptothecin conjugates may function as sequence-specific transcription terminators.

Recognition of Nucleosomes.

In eukaryotic cells, DNA is tightly packaged by compaction into chromatin, and changes in chromatin structure can alter the accessibility of specific sequences and affect components of the molecular machinery in the nucleus. The fundamental repeating unit of chromatin is the nucleosome, comprising a 20–80 bp DNA linker region and the nucleosome core particle (NCP)– roughly two tight superhelical turns of DNA (147 bp in length) wrapped around a disk of eight histone proteins. The ability of DNA-binding proteins to recognize their cognate sites in chromatin is restricted by the structure and

dynamics of nucleosomal DNA, and by the translational and rotational positioning of the histone octamer. Using six different hairpin polyamides, it was shown that sites on nucleosomal DNA facing away from the histone octamer, or even partially facing the octamers, are fully accessible.[80] Remarkably, one section of 14 consecutive base pairs— more than a full turn of the DNA helix—was accessible for high affinity polyamide binding. The only positions very poorly bound by polyamides were sites near the amino-terminal tails of histone H3 or histone H4. Removal of either tail allowed polyamides to bind, suggesting that the structure of the DNA and perhaps its rotational position are strongly influenced by the N-terminal tails of histone H3 and H4.[80]

Subsequently, the structures of three of these polyamide-NCP complexes were determined by X ray crystallography.[81] The histone octamer is unaffected by polyamide binding, but the nucleosomal DNA undergoes significant structural changes at the ligand binding sites and the adjacent regions. Significantly, distortions in DNA twist can propagate over long distances without disrupting histone-DNA contacts, giving a potential mechanistic rationale for the role of twist diffusion in nucleosome translocation. Although the three polyamides display very similar affinities for their binding sites in the α-satellite nucleosome particle, only the relatively non-specific polyamide ImPyPyPy-γ-PyPyPyPy-β-Dp inhibits temperature-induced nucleosome translocation.[81] This may indicate that ligand positioning is critical, such that a single properly placed polyamide would effectively block translocation; or that the small effects of a single bound ligand can be amplified, such that a combination of several different polyamides would block translocation.

Although polyamides can block transcription by targeting promoter elements,

they do not affect transcription when bound in the RNA coding regions of DNA.[82] Presumably, the strand melting required for RNA polymerase progression disrupts the minor groove and displaces polyamides. To investigate potential effects on transcription through a nucleosome, hairpin polyamides were targeted to sites on the nucleosome positioning sequence of the sea urchin 5S gene.[82] The two molecules that blocked heat-induced nucleosomal translocation also blocked transcription by T7 RNA polymerase. Each of these polyamides binds with high affinity to a single site in the nucleosome construct, potentially implying that placement is critical. Nonetheless, the positions of these sites are distinct from those occupied by the compound ImPyPyPy-γ-PyPyPyPy-β-Dp in the crystal structure.[81] Although the precise mechanisms involved in nucleosome repositioning remain in question, it appears that, in some cases, DNA can "roll" over the histones, and certain polyamides can act as chocks to prevent the DNA from moving.
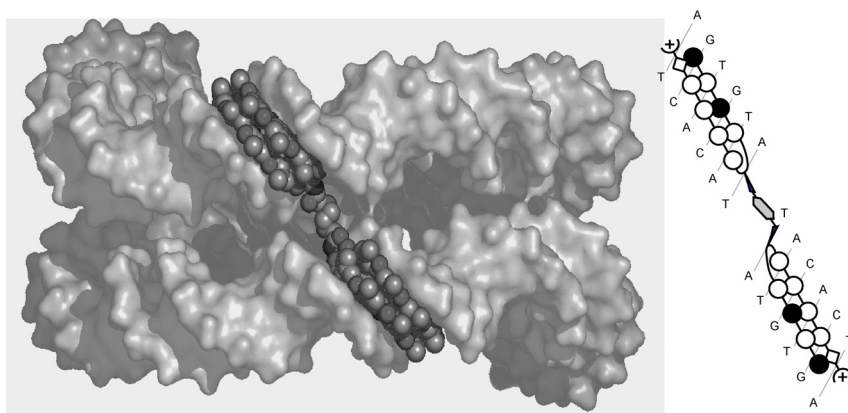
Figure 1.13. Left: Detail of the X ray crystal structure of the nucleosome core particle (NCP) with turn-to-turn dimer clamp **PW2** bound to a single supergroove. DNA helices run horizontally and are colored white. Polyamide bound in proximal supergroove is dark grey. At right is a schematic representation of binding, highlighting the DNA sequence.

One of the polyamide-NCP crystal structures showed a striking alignment of polyamides bound to sites almost 80 linear base pairs apart. Polyamides bound to these adjacent minor groove sites have their γ-turn moieties juxtaposed. Dimer polyamides

linked turn-to-turn bind to such "super groove" sites and serve as clamps, locking the DNA onto the nucleosome. (Figure 1.13).[83] This feature makes it possible to explore chromatin superhelix recognition. Such nucleosome targeting becomes relevant to biological systems, as the majority of cellular DNA is nucleosome-bound.

Nuclear Uptake.

DNA-binding polyamides can inhibit and influence a wide variety of protein-DNA interactions in solution, yet effectiveness in cell culture has proved to be dependent on cell type. A series of fluorescently labeled polyamides was prepared to analyze the intracellular distribution of these molecules in a panel of cell lines.[84] In cell types that had shown robust responses to polyamides, such as primary human T-cells, fluorescent polyamide-bodipy conjugates were observed to enter the nuclei of live cells.[84] However, in the majority of cell lines, polyamide-bodipy conjugates were excluded from the nucleus. Costaining with organelle-specific dyes indicates that polyamide-bodipy conjugates are often trapped in lysosomes and other cytoplasmic vesicles,[84] such that cells treated with polyamides can give a false nuclear signal upon fixing, even if they are washed extensively. Bashkin and coworkers have demonstrated that a polyamide-bodipy conjugate will traffic to the nucleus of a human cell line in the presence of verapamil, a p-glycoprotein inhibitor.[85]

Recently, a series of fluorescein-labeled polyamides were assayed for nuclear uptake against a panel of live mammalian cells.[86, 87] In some cases, small changes, such as the removal of a β-Ala residue at the C-terminus of a polyamide dramatically enhanced nuclear localization (Figure 1.14). Nuclear uptake of tested polyamide-fluorescein
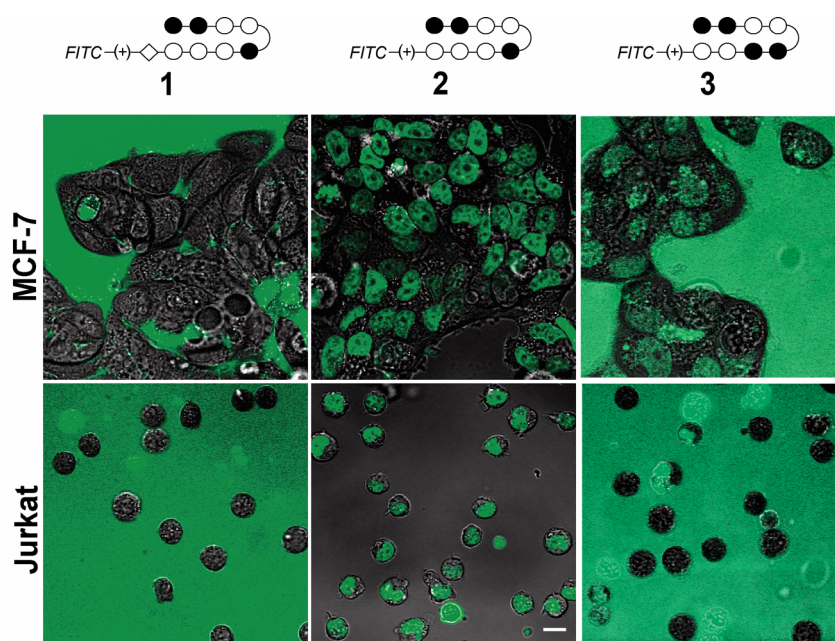
Figure 1.14. Nuclear uptake of hairpin polyamides. Representative confocal microscopy images of fluorescein-labeled polyamides in MCF-7 and Jurkat cells. Polyamide **1** exhibits poor uptake and is excluded from the nuclei of both cell types. Removal of the C-terminal β-Ala residue results in a polyamide (**2**) with excellent uptake properties, localizing to the nuclei of both cell types. Polyamide **3** differing from **2** by a single pyrrole to imidazole substitution, localizes to MCF-7 nuclei less strongly than does **2**, and **3** is completely excluded from Jurkat cells.

conjugates is an energy-dependent process. HeLa cells grown in energy inhibitory medium (supplemented with 2-deoxyglucose and sodium azide) displayed little to no discernable nuclear staining when treated with a fluorescein-labeled polyamide, while the same cells grown in normal medium showed clear nuclear staining. Washing of the inhibitory medium, and replacement with normal medium (supplemented with additional polyamide) resulted in nuclear staining.

While there are currently no general rules for cellular uptake of polyamides, determinants such as polyamide size, imidazole content, structure and attachment point of the fluorescent dye, and structure of the "tail" are important for nuclear localization. Each cell line possess a unique uptake profile such that choices of specific cell lines and compound architectures will be critical for future biological experiments.

Scope of This Work.

The work presented in this thesis is mainly focused towards using the target DNA to template the formation of higher-order polyamide structures with improved binding properties. As has been noted in this introduction, hairpin dimers, while targeting long sequences of DNA with high affinity and specificity, are unable to translocate to the nuclei of living cells, and thus cannot be used for experiments in organisms. We report our efforts towards forming these higher-order dimeric polyamides from smaller, potentially cell-permeable starting materials. Chapter 2 of this thesis is concerned with turn-to-tail templated reactions of duplex DNA. Chapter 3 of this thesis is concerned with turn-to-turn templated reactions both on duplex DNA and across NCP supergrooves. Chapter 4 of this thesis presents our efforts towards non-covalent interactions between polyamide strands bound in the minor groove.

This thesis also contains three appendices. While much of polyamide research has focused on biological applications such as gene regulation, polyamides possess amazing DNA-binding properties that make them well suited for use as molecular biology tools. Appendix I of this thesis details our efforts towards the use of polyamide-biotin conjugates for the capture and purification of DNA sequences from fragmented genomes. Appendix III of this thesis details our efforts towards the use of polyamides as specificity agents for structure determination of non-sequence-specific DNA enzymes.

Finally, Appendix II of this thesis examines the DNA-binding properties of tail-less polyamides synthesized on oxime resin.

References

1. Arcamone, F.; Nicolell, V; Penco, S.; Orezzi, P.; Pirelli, A., *Nature*, **1964,** *203*, 1064–1066.

2. Zimmer, C.; Wahnert, U., *Progress in Biophysics & Molecular Biology*, **1986,** *47*, 31–112.

3. Schuhman, E.; Haupt, I.; Thrum, H.; Taubenec.U; May, U., *Zeitschrift Fur Allgemeine Mikrobiologie*, **1974,** *14*, 321–327.

4. Ginsburg, H.; Nissani, E.; Krugliak, M.; Williamson, D. H., *Molecular and Biochemical Parasitology*, **1993,** *58*, 7–15.

5. Thrum, H.; Haupt, I.; Bradler, G.; Zimmer, C. G.; Reinert, K. E., *Antimicrobial and Antineoplastic Chemotherapy*. Czech. Med. Press: Prague, 1972; Vol. 1, p. 819.

6. Verini, M. A.; Ghione, M., *Chemotherapy*, **1964,** *9*, 144–148.

7. Mars, G.; Regoli, U., *Clin. Ter.*, **1968,** *30*, 573–580.

8. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E., *Proceedings of the National Academy of Sciences of the United States of America*, **1985,** *82*, 1376–1380.

9. Pelton, J. G.; Wemmer, D. E., *Proceedings of the National Academy of Sciences of the United States of America*, **1989,** *86*, 5723–5727.

10. Coll, M.; Frederick, C. A.; Wang, A. H. J.; Rich, A., *Proceedings of the National Academy of Sciences of the United States of America*, **1987,** *84*, 8385–8389.

11. Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology*, **2003,** *13*, 284–299.

12. Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2001,** *9*, 2215–2235.

13. Wade, W. S.; Mrksich, M.; Dervan, P. B., *Journal of the American Chemical Society*, **1992,** *114*, 8783–8794.

14. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nature Structural Biology*, **1998,** *5*, 104–109.

15. Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J., *Biochemistry*, **1999,** *38*, 2143–2151.

16. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Nature*, **1998,** *391*, 468–471.

17. 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.

18. Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., *Journal of Molecular Biology*, **2000,** *295*, 557–567.

19. Wellenzohn, B.; Loferer, M. J.; Trieb, M.; Rauch, C.; Winger, R. H.; Mayer, E.; Liedl, K. R., *Journal of the American Chemical Society*, **2003,** *125*, 1088–1095.

20. Mrksich, M.; Parks, M. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1994,** *116*, 7983–7988.

21. Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature*, **1996,** *382*, 559–561.

22. deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society*, **1997,** *119*, 7909–7916.

23. White, S.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1997,** *119*, 8756–8765.

24. Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1999,** *121*, 1113–1120.

25. Hawkins, C. A.; de Clairac, R. P.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society*, **2000,** *122*, 5235–5243.

26. Herman, D. M.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1998,** *120*, 1382–1391.

27. Dervan, P. B.; Urbach, A. R., In *Essays in Contemporary Chemistry*, Quinkert, G.; Kisakurek, M. V., Eds. Verlag Helv. Chim. Acta: 2001; p. 327–339.

28. Urbach, A. R. 1:1 Motif for DNA REcognition by β–Alanine-Linked Polyamides. California Institute of Technology, Pasadena, 2002.

29. Woods, C. R.; Ishii, T.; Wu, B.; Bair, K. W.; Boger, D. L., *Journal of the American Chemical Society*, **2002,** *124*, 2148–2152.

30. Kelly, J. J.; Baird, E. E.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **1996,** *93*, 6981–6985.

31. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1998,** *120*, 6219–6226.

32. Wang, C. C. C.; Ellervik, U.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2001,** *9*, 653–657.

33. Herman, D. M.; Baird, E. E.; Dervan, P. B., *Chemistry-a European Journal*, **1999,** *5*, 975–983.

34. Kers, I.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 3339–3349.

35. Weyermann, P.; Dervan, P. B., *Journal of the American Chemical Society*, **2002,** *124*, 6872–6878.

36. Maeshima, K.; Janssen, S.; Laemmli, U. K., *Embo Journal*, **2001,** *20*, 3218–3228.

37. Greenberg, W. A.; Baird, E. E.; Dervan, P. B., *Chemistry-a European Journal*, **1998,** *4*, 796–805.

38. Olenyuk, B.; Jitianu, C.; Dervan, P. B., *Journal of the American Chemical Society*, **2003,** *125*, 4741–4751.

39. Heckel, A.; Dervan, P. B., *Chemistry-a European Journal*, **2003,** *9*, 3353–3366.

40. Hunter, C. A., *Journal of Molecular Biology*, **1993,** *230*, 1025–1054.

41. Urbach, A. R.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2001,** *98*, 4343–4348.

42. Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., *Journal of Molecular Biology*, **2002,** *320*, 55–71.

43. Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B., *Helvetica Chimica Acta*, **2002,** *85*, 4485–4517.

44. Janssen, S.; Cuvier, O.; Muller, M.; Laemmli, U. K., *Molecular Cell*, **2000,** *6*, 1013–1024.

45. Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6141–6146.

46. Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Organic Letters*, **2001,** *3*, 1201–1203.

47. Fattori, D.; Kinzel, O.; Ingallinella, P.; Bianchi, E.; Pessi, A., *Bioorganic & Medicinal Chemistry Letters*, **2002,** *12*, 1143–1147.

48. Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 2767–2774.

49. Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2003,** *11*, 4333–4340.

50. Minehan, T. G.; Gottwald, K.; Dervan, P. B., *Helvetica Chimica Acta*, **2000,** *83*, 2197–2213.

51. Briehn, C. A.; Weyermann, P.; Dervan, P. B., *Chemistry-a European Journal*, **2003,** *9*, 2110–2122.

52. Renneberg, D.; Dervan, P. B., *Journal of the American Chemical Society*, **2003,** *125*, 5707–5716.

53. Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature*, **1997,** *387*, 202–205.

54. Darnell, J. E., *Nature Reviews Cancer*, **2002,** *2*, 740–749.

55. Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Gottesfeld, J. M., *Journal of Molecular Biology*, **1997,** *274*, 439–445.

56. Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **1998,** *95*, 12890–12895.

57. Ehley, J. A.; Melander, C.; Herman, D.; Baird, E. E.; Ferguson, H. A.; Goodrich, J. A.; Dervan, P. B.; Gottesfeld, J. M., *Molecular and Cellular Biology*, **2002,** *22*, 1723–1733.

58. McBryant, S. J.; Baird, E. E.; Trauger, J. W.; Dervan, P. B.; Gottesfeld, J. M., *Journal of Molecular Biology*, **1999,** *286*, 973–981.

59. Winston, R. L.; Ehley, J. A.; Baird, E. E.; Dervan, P. B.; Gottesfeld, J. M., *Biochemistry*, **2000,** *39*, 9092–9098.

60. Lenzmeier, B. A.; Baird, E. E.; Dervan, P. B.; Nyborg, J. K., *Journal of Molecular Biology*, **1999,** *291*, 731–744.

61. Simon, H.; Kittler, L.; Baird, E.; Dervan, P.; Zimmer, C., *Febs Letters*, **2000,** *471*, 173–176.

62. Wurtz, N. R.; Pomerantz, J. L.; Baltimore, D.; Dervan, P. B., *Biochemistry*, **2002,** *41*, 7604–7609.

63. Yang, F.; Belitsky, J. M.; Villanueva, R. A.; Dervan, P. B.; Roth, M. J., *Biochemistry*, **2003,** *42*, 6249–6258.

64. Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Dervano, P. B.; Graves, B. J.; Gottesfeld, J. M., *Journal of Biological Chemistry*, **1999,** *274*, 12765–12773.

65. Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A., *Journal of Biological Chemistry*, **2000,** *275*, 24246–24254.

66. Schaal, T. D.; Mallet, W. G.; McMinn, D. L.; Nam, V. N.; Sopko, M. M.; John, S.; Parekh, B. S., *Nucleic Acids Research*, **2003,** *31*, 1282–1291.

67. Bremer, R. E.; Baird, E. E.; Dervan, P. B., *Chemistry & Biology*, **1998,** *5*, 119–133.

68. Bremer, R. E.; Wurtz, N. R.; Szewczyk, J. W.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2001,** *9*, 2093–2103.

69. Fechter, E. J.; Dervan, P. B., *Journal of the American Chemical Society*, **2003,** *125*, 8476–8485.

70. Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Ghazal, P.; Dervan, P. B.; Gottesfeld, J. M., *Biochemistry*, **1999,** *38*, 10801–10807.

71. Coull, J. J.; He, G. C.; Melander, C.; Rucker, V. C.; Dervan, P. B.; Margolis, D. M., *Journal of Virology*, **2002,** *76*, 12349–12354.

72. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2000,** *97*, 3930–3935.

73. Ansari, A. Z.; Mapp, A. K.; Nguyen, D. H.; Dervan, P. B.; Ptashne, M., *Chemistry & Biology*, **2001,** *8*, 583–592.

74. Arora, P. S.; Ansari, A. Z.; Best, T. P.; Ptashne, M.; Dervan, P. B., *Journal of the American Chemical Society*, **2002,** *124*, 13067–13071.

75. Passner, J. M.; Ryoo, H. D.; Shen, L. Y.; Mann, R. S.; Aggarwal, A. K., *Nature*, **1999,** *397*, 714–719.

76. Arndt, H. D.; Hauschild, K. E.; Sullivan, D. P.; Lake, K.; Dervan, P. B.; Ansari, A. Z., *Journal of the American Chemical Society*, **2003,** *125*, 13322–13323.

77. Wang, C. C. C.; Dervan, P. B., *Journal of the American Chemical Society*, **2001,** *123*, 8657–8661.

78. Hsiang, Y. H.; Hertzberg, R.; Hecht, S.; Liu, L. F., *Journal of Biological Chemistry*, **1985,** *260*, 4873–4878.

79. Bendixen, C.; Thomsen, B.; Alsner, J.; Westergaard, O., *Biochemistry*, **1990,** *29*, 5613–5619.

80. Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B., *Journal of Molecular Biology*, **2001,** *309*, 615–629.

81. Suto, R. K.; Edayathumangalam, R. S.; White, C. L.; Melander, C.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Journal of Molecular Biology*, **2003,** *326*, 371–380.

82. Gottesfeld, J. M.; Belitsky, J. M.; Melander, C.; Dervan, P. B.; Luger, K., *Journal of Molecular Biology*, **2002,** *321*, 249–263.

83. Edayathumangalam, R. S.; Weyermann, P.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Proceedings of the National Academy of Sciences of the United States of America*, **2004,** *101*, 6864–6869.

84. Belitsky, J. M.; Leslie, S. J.; Arora, P. S.; Beerman, T. A.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 3313–3318.

85. Crowley, K. S.; Phillion, D. P.; Woodard, S. S.; Schweitzer, B. A.; Singh, M.; Shabany, H.; Burnette, B.; Hippenmeyer, P.; Heitmeier, M.; Bashkin, J. K., *Bioorganic & Medicinal Chemistry Letters*, **2003,** *13*, 1565–1570.

86. Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2003,** *100*, 12063–12068.

87. Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B., *Nucleic Acids Research*, **2004,** *32*, 2802–2818.

Chapter 2

Turn-to-Tail Dimerizations of Hairpin Polyamides on Duplex DNA Templates

(Poulin-Kerstien, A. T.; Dervan, P. B. "DNA-Templated Dimerizations of Hairpin Polyamides." *J. Am. Chem. Soc.* **2003**, *125*, 15811–15821)

Abstract

Double-helical DNA accelerates the rate of ligation of two six-ring hairpin polyamides, which bind adjacent sites in the minor groove via a 1,3-dipolar cycloaddition to form a tandem dimer. The rate of the templated reaction is dependent on DNA sequence as well as on the distance between the hairpin binding sites. The tandem-dimer product of the DNA-templated reaction has improved binding properties with respect to the smaller hairpin fragments. Since cell and nuclear uptake of DNA-binding polyamides will likely be dependent on size, this is a minimum first step toward the design of self-assembling small gene-regulating fragments to produce molecules of increasing complexity with more specific genomic targeting capabilities.

Introduction.

Small molecules that bind a large repertoire of DNA sequences and modulate transcription could be useful in biology and medicine.[1, 2] Polyamides comprised of three aromatic amino acids, N-methylpyrrole (Py), N-methylimidazole (Im), and N-methyl-3-hydroxypyrrole (Hp), distinguish the four Watson-Crick base pairs by a set of pairing rules.[3] Connecting the two antiparallel strands of aromatic amino acids with a γ-aminobutyric acid (γ) creates a hairpin motif capable of binding to match DNA with increased affinity and sequence specificity.[4-6] For applications in gene regulation within biological systems, binding-site size may be critical because longer sequences should occur less frequently in a gigabase-sized genome. For this reason, the design of ligands capable of targeting >10 base pairs (bp) of DNA remains an important goal in the area of polyamide design.[3, 7-9] An ideal DNA-binding polyamide used for gene regulation must possess several properties: high affinity to DNA, such that it can compete with cellular DNA-binding proteins; specificity, to distinguish its targeted match site from mismatch sites; and favorable cell and nuclear uptake properties in order to reach its targeted DNA on nuclear chromatin.

DNA-binding eight-ring hairpin polyamides possess excellent affinity and sequence specificity, but they target only 6 bp.[3, 7-9] Various polyamide motifs have been designed to target longer sequences. High resolution X ray studies reveal that polyamides containing more than five contiguous aromatic ring pairings are over curved with respect to the DNA helix, which results in a loss of the hydrogen bonds and van der Waals interactions responsible for binding affinity.[10, 11] Replacement of one or more internal Py carboxamides with more flexible β-alanine (β) residues relaxes the polyamide

curvature and allows longer hairpin polyamides to bind DNA with restored affinity.[12-14]

These types of flexible motifs have been used to target up to 16 base pairs of DNA.[12-14]

However, polyamides containing internal β-alanines are also able to bind in 1:1 ligand-DNA stoichiometries, thereby decreasing their specificity.[15, 16] Another approach to increase polyamide binding-site size has been to covalently link two hairpin modules to form hairpin dimers.   Dimers linked both "turn-to-tail" and "turn-to-turn" have excellent affinity and specificity to DNA sequences up to 10 bp in length.[17-19] Though likely satisfying the DNA-binding criteria to target unique sequences within large genomic DNA, hairpin dimers do not possess the favorable cell and nuclear uptake properties of smaller hairpins, presumably due to size and shape.[20, 21]

Many DNA-binding transcription factors and complexes rely on dimerization or multimerization of DNA recognition elements that each occupy 4–6 base pairs and target unique contiguous sites in genomic DNA.[22, 23] This cooperative, combinatorial association of gene-regulatory proteins may be a useful strategy to overcome the kinetic problems associated with finding long contiguous sequences of DNA. As polyamides are designed to target longer sequences of DNA, a similar kinetic barrier may be encountered.

Single-strand DNA has long been known to template chemical reactions by bringing reactive functionalities in close proximity.[24-45] Annealing of two adjacent complementary single strands of DNA brings the reactive groups together. Remarkably, Liu and coworkers report that single-stranded DNA can template various chemical reactions in systems in which 1–30 nucleobases of single-stranded DNA separate the template site from the reactive site, with little observable change in reactivity.[26, 30-33] The

single-stranded template DNA serves to coordinate chemical reactivity and is not a desired part of the final product.[26, 30-33]

In this report we explore *double-stranded DNA-templated reactions* combining hairpin modules in the minor groove with the overall goal of producing larger polyamides capable of targeting longer sequences.[46] Two different six-ring hairpin polyamides were designed such that when their match sites are adjacent on the DNA a thermal reaction at 37 °C (pH 7.0) would afford a covalent bond between the hairpin modules, forming a tandem-dimer structure *in situ* (Figure 2.1). The tandem-hairpin dimer product should have improved DNA-binding properties over the smaller hairpin subunits. This type of scheme differs from hybridization-based DNA-templated chemistry in that we are using the double-helical structure of DNA to template covalent bond formation. In this method
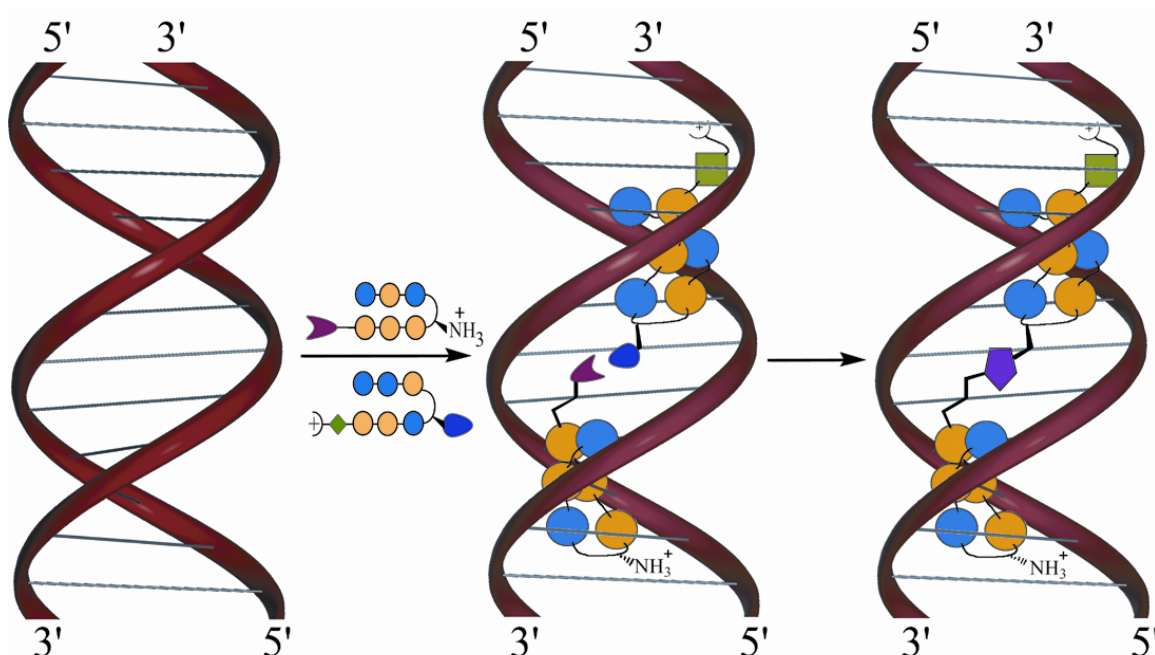
Figure 2.1. Schematic model of DNA-templated tandem hairpin formation. Polyamides functionalized with complementary reactive groups (red and blue shapes) bind to contiguous match sites on DNA. The reactive groups are placed in close proximity, causing them to form a covalent bond, linking the hairpins (purple pentagon).

pyrrole-imidazole polyamides read by "pairing rules" the unique hydrogen bonding pattern presented by the edges of the Watson-Crick base pairs in the minor groove of DNA. The DNA binding hairpin polyamide modules become part of a ligand-DNA supramolecular complex containing four oligomers: two DNA strands and two juxtaposed polyamides. *The sequence information encoded in the DNA base pairs becomes encoded in the polyamide product molecule.*[46]

Results.

Chemical Ligation Reaction. The reaction chosen to ligate the DNA-binding hairpins together is the 1,3-dipolar cycloaddition reaction of an acetylene and azide.[47-52] The active site of acetylcholine esterase has been shown to template the formation of a high-affinity inhibitor from azide- and alkyne-functionalized building blocks utilizing the Hüisgen 1,3-cycloaddition.[49] The thermal cycloaddition reaction requires no cofactors and tolerates reaction conditions similar to those inside a cell.[47-52] The simultaneous binding of two ligands on adjacent addressable sites should accelerate the reaction that connects them. To create unique points of reactivity in the minor groove, an acetylene on the γ-turn of one hairpin would be placed proximal to the azide on the C-terminus of an adjacent hairpin in tandem orientations affording a 1,2,3-triazole in the linker of a tandem hairpin dimer product (Fig. 2.2).
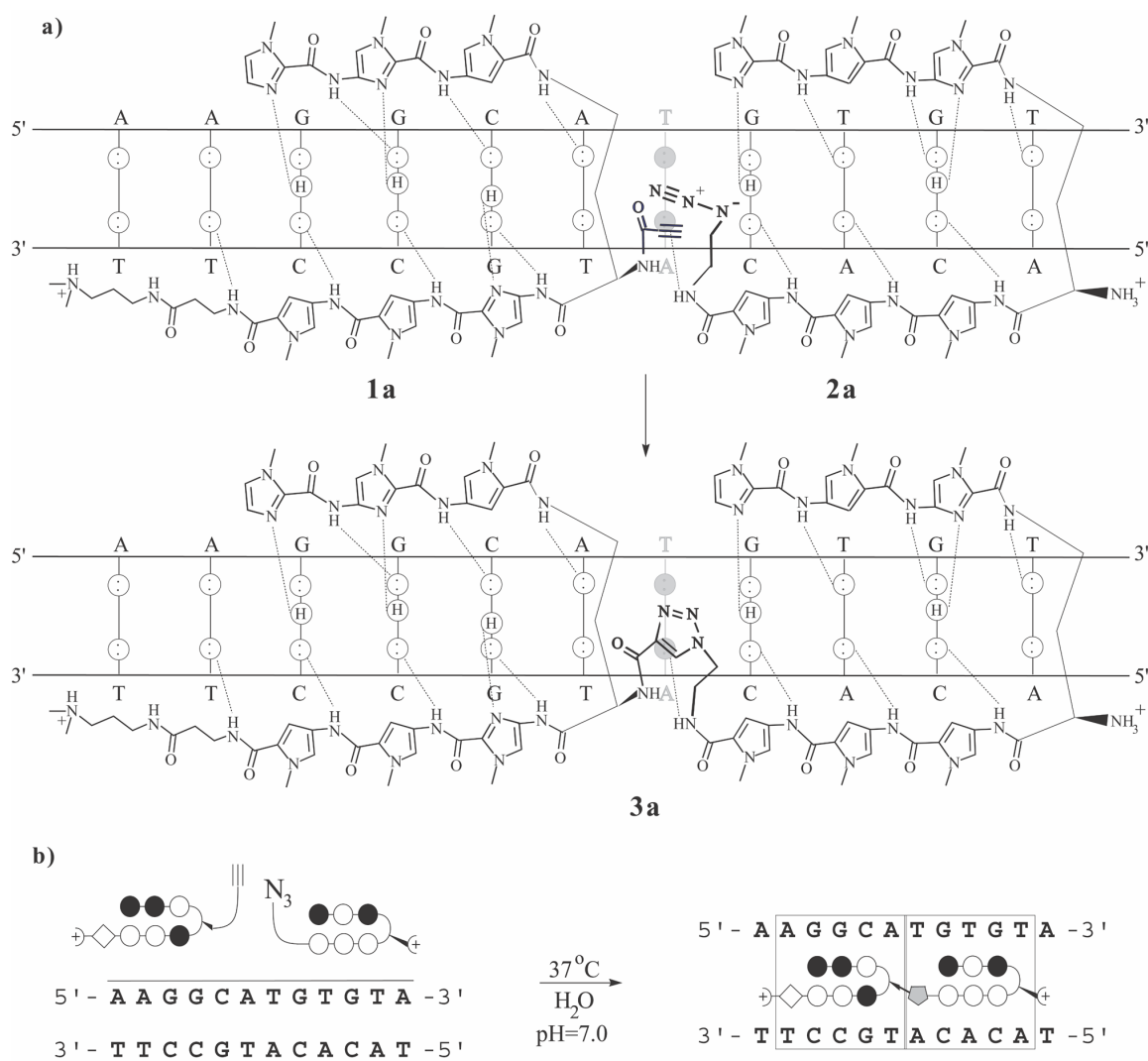
Figure 2.2. a) Hydrogen bonding model of the hairpin-DNA complex, **1a**, **2a**, at the 10-bp match (zero intervening bp) site 5'-AGGCATGTGT-3' (top), and reaction product tandem **3a** complexed with DNA (bottom). Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines. b) Schematic model of alkyne **1a** and azide **2a** binding to their match sites and forming **3a**, which recognizes the entire 10-bp binding site. Im and Py residues are represented by filled and open circles, respectively. The β-residue is represented by a diamond. The triazole linker is represented as a pentagon. Individual binding sites for the starting hairpin polyamides (**1a** and **2a**) are boxed.

Hairpin polyamide design. For the design of the hairpin polyamide reactants the flexibility and distance of the linker that connects the reactive alkyne and azide functionalities to the DNA-binding polyamides were chosen based on model building. With regard to reaction times at 37 °C, we anticipated that the inherent reactivity of the

starting materials could be tuned to afford templated intramolecular reaction half-lives on the order of hours, rather than seconds or weeks. For example, alkynes possessing an electron-withdrawing group, such as an amide or carboxylic ester, are more reactive in dipolar cycloadditions than their alkyl counterparts (up to 10^5 increase in reaction rates).[47, 51, 52] Alkyne-functionalized polyamides ImImPy-(R)-γ^{NH}[COC≡CH]-ImPyPy-β-Dp (**1a**), ImImPy-(R)-γ^{NH}[CO(CH$_2$)$_2$C≡CH]-ImPyPy-β-Dp (**1b**) were synthesized (Dp = 3-(Dimethylamino)propylamine, γ = (R)-2,4-diaminobutyric acid) with two different linker lengths; **1a** possessing the rigid alkynyl amidate functionality and **1b** possessing the more flexible alkyl alkyne. Our criteria for screening linkers for the DNA-templated reaction was to select reactive partners that do not react in solution at μM concentration (37 °C) but react in a reasonable time (hours) on DNA. Azide-functionalized polyamides ImPyIm-(R)-γ^{NH_2}-PyPyPy-(CH$_2$)$_2$-N$_3$ (**2a**) and ImPyIm-(R)-γ^{NH_2}-PyPyPy-(CH$_2$)$_3$-N$_3$ (**2b**) were synthesized which contain either an ethyl (for **2a**) or propyl (for **2b**) linker connecting the azide moiety to the C-terminus. These molecules were chosen because the 1,2,3-triazole linker (Tr) formed in the pairing of either **2a–b** with **1a** best approximated the length of the 5-aminovaleric acid linker found in our previous studies to be optimal for "turn-to-tail" tandem affinity and specificity to a 10 bp binding site.[18]

Precursor **4** was synthesized on resin using Boc chemistry and was liberated by aminolysis with neat Dp (Figure 2.3).[53] Six-ring hairpins **1a** and **1b** were completed by the installation of the alkynyl functionality on the α-amine of the γ-turn residue using either propiolic acid (for **1a**) or 1-pentynoic acid (for **1b**) and DCC coupling conditions according to modified literature procedures.[54] Azide-functionalized polyamide

precursors **6a–b** were synthesized on Kaiser oxime resin[55] and were liberated from solid

support by aminolysis with either 2- or 3-aminoethanol (**6a** and **6b**, respectively).
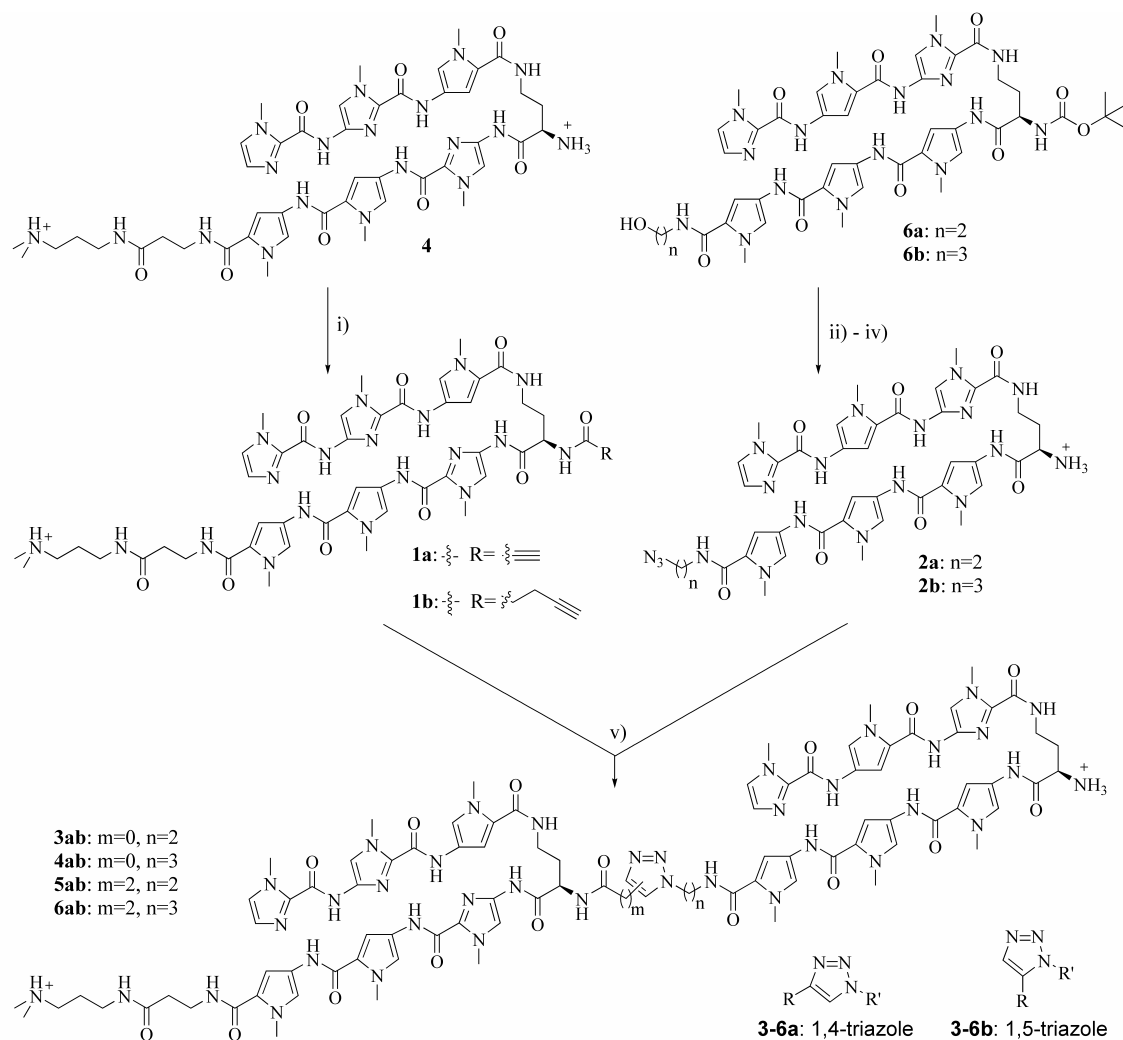
Figure 2.3. Synthesis of hairpins **1a–b** and **2a–b** and tandem hairpin dimers **3ab–6ab**: i) propiolic acid (for **1a**) or 5-pentynoic acid (for **1b**), DCC, DMF/CH$_3$CN 1:1, 0 °C, 6 h; ii) DIEA, CH$_2$Cl$_2$, 0 °C, 15 min, then MsCl, rt, 2 h; iii) NaN$_3$, DMF, 70 °C, 12 h; (iv) CH$_2$Cl$_2$-TFA 1:1, 15 min; (v) neat, 60 °C, 5–14 days.

Mesylation of the resulting alcohols, followed by displacement with sodium azide

resulted in the appropriately functionalized C-termini.[56] Compounds **2a** and **2b** were

obtained following Boc-deprotection of the α–amino group of the γ-turn residue with

50% TFA-CH$_2$Cl$_2$.

Authentic samples of the expected tandem products ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_2$-Tr-(OC)]^{HN}γ-ImPyPy-β-Dp   (**3ab**),   ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_3$-Tr-(OC)]^{HN}γ-ImPyPy-β-Dp   (**4ab**),   ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_2$-Tr-(CH$_2$)$_2$(OC)]^{HN}γ-ImPyPy-β-Dp (**5ab**),   and   ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_3$-Tr-(CH$_2$)$_2$(OC)]^{HN}γ-ImPyPy-β-Dp   (**6ab**)   were synthesized by heating dry, powdered mixtures of **1a-b** and **2a-b** at 60 °C for 5 days.

1,3-dipolar cycloadditions between azides and alkynes are known to produce both 1,4- and 1,5-substituted triazole ring products. In the case of alkyl substituted reactants, the stereoelectronics of the reaction pathways leading to each regioisomer are approximately equal, leading to equal ratios of the two regioisomers. When the alkynyl reactant is substituted with an electron withdrawing group, the pathway leading to the 1,4-regioisomer becomes favored, producing this isomer as the major product.[47, 51, 52] The thermal reaction between activated alkyne **1a** and azide **2a** favors the 1,4-regioisomeric product **3a** over the 1,5-regioisomeric product **3b** by a ratio of 20:1.[57] Regioisomers of substituted triazole rings are known to have distinct chemical shifts for the lone aromatic proton, and this distinction is the basis for our assignment of regioisomers.[58] Similarly, the reaction between the activated alkyne **1a** and the longer azide **2b** to form **4ab** gives a product ratio of 20:1, presumably **4a**:**4b**. When the unactivated alkyl alkyne **1b** is paired with either azide **2a** or **2b** to form **5ab** and **6ab**, respectively, each thermal coupling yields the two expected regioisomers in a ratio of 1:1. We anticipate the DNA-templated cycloaddition in the minor groove might produce different regioisomeric ratios due to steric constraints within the polyamide/DNA complex.

DNA-templated tandem formation.

Reactivity and hairpin binding site separation preference (Table 2.1). A key issue in evaluating the DNA-templated reaction was to determine if the cycloaddition reaction was accelerated in the presence of match DNA and in reasonable yields. In addition, we anticipated that reactions would be sensitive to the distance separating the hairpin-binding modules.

Table 2.1. Relative pseudozero-order rate constants (s^{-1}) for DNA-templated tandem formation when the reaction is performed with the hairpin binding sites separated by zero, one, or two intervening base pairs.[a-c]

	5'-AGGCA**T**GTGT-3' **A**	5'-AGGCA**A**TGTGT-3' **B**	5'-AGGCA**AA**TGTGT-3' **C**	No DNA
1a + 2a	16290 (±54)	≤ 1	≤ 1	≤ 1
1a + 2b	13040 (±42)	≤ 1	≤ 1	≤ 1
1b + 2a	200 (±6)	≤ 1	≤ 1	≤ 1
1b + 2b	132 (±1)	21 (±1)	≤ 1	≤ 1

[a] The reported rates are normalized with respect to the nontemplated reaction between **1a** and **2a** at 1 μM, and are the average values obtained from three kinetics experiments, with the error for each data set indicated in parentheses. [b] The assays were carried out at 37 °C at pH 7.0 in the presence of 3 mM Tris-HCl, 3 mM KCl, 3 mM $MgCl_2$, and 1.7 mM $CaCl_2$, 1 μM each polyamide, 1μM DNA. [c] Rate data were taken from the linear phase of product formation (four time points over 5.25 hours), except for experiments with **1b**, which were taken from three time points over two weeks.

Tandem-hairpin polyamides have previously been shown to bind to both 10- and 11-bp sites, with zero and one base pairs separating the hairpin binding sites, respectively. To assess the appropriate polyamide separation distance for the DNA-templated tandem-forming cycloaddition, the duplex DNA templates 5'-GGGGT**AGGCATGTGT**AGGGG-3' (**A**), 5'-GGGGT**AGGCAA** **TGTGT**AGGGG-3' (**B**) and 5'-GGGGT**AGGCA**AA**TGTGT**AGGGG-3' (**C**) were synthesized (Figure 2.4a).

Each duplex contains five bp match sites for the two hairpin polyamides **1(a–b)** and **2(a–b)** separated by zero, one, or two base pairs, respectively.

Reactions were performed with equal concentrations of each hairpin polyamide and DNA (2 mM Tris-HCl, 2 mM KCl, 2 mM MgCl$_2$, 1 mM CaCl$_2$, pH 7.0, 37 °C). Analytical reverse-phase HPLC was used to monitor the cycloaddition reactions. MALDI-TOF MS was used to verify HPLC assignments. Experiments were carried out at 1 μM concentrations of DNA and hairpin polyamides.

When any pair of hairpin polyamides (**1a** + **2a**, **1a** + **2b**, **1b** + **2a**, **1b** + **2b**) is combined in solution at 1 μM concentration of each polyamide in the absence of DNA, no tandem product is observed after two weeks at 37 °C (pH = 7.0). The detection limit is approximately 0.1% of the total starting material. Thus, all tandem-forming cycloadditions proceed in less than 0.1% yield after 2 weeks in the absence of DNA.

When alkyne **1a** and ethyl azide **2a** are incubated with duplex **A** (zero intervening bases) at 1 μM concentrations, tandem product **3ab** is formed in detectable amounts after 45 minutes (37 °C, pH = 7.0). Quantitation of HPLC traces shows that **3ab** is formed in approximately 35% yield in 5 hours. We observe that the reaction does not progress significantly after this, exhibiting 43% product formation after 24 hours. When **1a** and **2a** are incubated with either of the longer duplexes **B** or **C** (1 and 2 intervening base pairs, respectively), tandem product **3ab** is not detected after 2 weeks. When **1a** and propyl azide **2b** are incubated with duplex **A** (zero intervening bases) at 1 μM, tandem product **4ab** is formed 20% more slowly than **3ab** from **1a** and **2a** on the same duplex. **1a** and **2b** produce no observable product formation on either duplex **B** or **C**. When less reactive alkyne **1b** and ethyl azide **2a** are incubated with duplex **A** (zero intervening

bases), **5ab** is formed in 20% yield after 2 weeks. When either one or two intervening

bases is present (duplex **B** and **C**, respectively), **5ab** is not detected. Product **6ab** is

formed on duplex **A** (zero intervening bases) from the long, flexible alkyne **1b** and propyl

azide **2b** in 13% yield in 2 weeks. These long linkers are able to span an additional bp,

with product **6ab** forming from this pairing on duplex **B** in 6% yield in 2 weeks.

Template mismatch tolerance: 10bp binding site (Table 2.2)**.**   To assess if the

cycloaddition is sequence-specific with respect to the template, the duplexes

Table 2.2. Relative pseudozero-order rate constants (s^{-1}) for DNA-templated tandem formation of **3ab** from **1a** and **2a**. This data was collected on the duplex template (zero intervening bp) to probe tandem formation when there are mismatches under one or both hairpin sites.[a-d]

Concentration	5'-AGGCATGTCT-3' A	5'-AGGCATGTCT-3' D	5'-AGGGATGTCT-3' E	No DNA
1.0 μM	16290 (±51) [42.7% (±0.7)]	10441 (±210) [28.2% (±0.2)]	1540 (±30) [16.3% (±0.2)]	1 [≤0.1%]
750 nM	17130 (±92) [47.3% (±0.7)]	11033 (±141) [30.4% (±0.4)]	793 (±3) [7% (±1)]	1 [≤0.1%]
500 nM	15420 (±77) [41% (±2)]	8927 (±327) [26% (±1)]	1050 (±22) [7.7% (±0.1)]	1 [≤0.1%]

[a] The reported rate constants are normalized with respect to that of **1a** and **2a** with no DNA template and are the average values obtained from three kinetics experiments, with the error in each data set indicated in parentheses. [b] The assays were carried out at 37 °C at pH 7.0 in the presence of 3 mM Tris-HCl, 3 mM KCl, 3 mM $MgCl_2$, and 1.7 mM $CaCl_2$, and the listed concentration of each polyamide and DNA. [c] Rate data was taken from the linear phase of product formation (4 time points over 5.25 hours). [d] In brackets is the amount of tandem **3ab** formed after 24-hour incubation (expressed as a percentage with 100% being complete conversion to product).

corresponding   to   5'-GGGGT**AGGCATGTCT**AGGGG-3'   (**D**)   and   5'-

GGGGT**AGGGATGTCT**AGGGG-3' (**E**) were synthesized (Figure 2.4a).   These

duplexes contain a single bp mismatch under one hairpin binding site (**D**) and a single bp

mismatch under each of the two hairpin polyamide binding sites (**E**) for the designed site

with zero intervening bp. This site was chosen because hairpin site-separation preference results indicated the 10 bp site to be optimal for tandem formation.

Because **1a** and **2a** showed the best DNA-templated reaction rate and yield, this pair was chosen for mismatch tolerance studies. Tandem product **3ab** is formed on double-stranded DNA template **D** 0.65 times as fast as on duplex **A** (zero intervening base pairs, match site). Product formation goes to only 28% completion on duplex **D** after 24 hours. When an additional mismatch is introduced (duplex **E**), **1a** and **2a** react to form **3ab** 0.029 times as fast as at the match site (duplex **A**). Product formation reaches only 16% after 24 hours (Figure 2.4b).

Reaction order and product verification. While non-templated couplings of **1b** with either **2a** or **2b** yield **5ab** and **6ab**, respectively, in equal regioisomeric ratios, when the reaction between **1b** and **2a** is templated on duplex **A** (zero intervening bp), a single regioisomer is formed. Likewise, when duplex **B** (one intervening bp) is used to template the formation of **6ab** from **1b** and **2b**, the reaction produces only a single regioisomer. When the reaction between **1b** and **2b** is templated by duplex **A** (zero intervening bp), two regioisomers are produced in a ratio of 3:1. When activated alkyne **1a** is paired with either **2a** or **2b** on template **A**, a single product isomer is produced.

When **1a** and **2a** are assayed on duplex DNA templates **A**, **D**, and **E** at either 750 nM or 500 nM concentrations, the relative rates of tandem formation are similar to those observed at 1 µM concentrations (Figure 2.4c). After 24 hours, the tandem-forming reaction is 47% complete at 750 nM and 41% complete at 500 nM. The reaction mixture of **1a** and **2a** at 1 µM concentration on DNA template **A** was taken at 8.25 hours and an
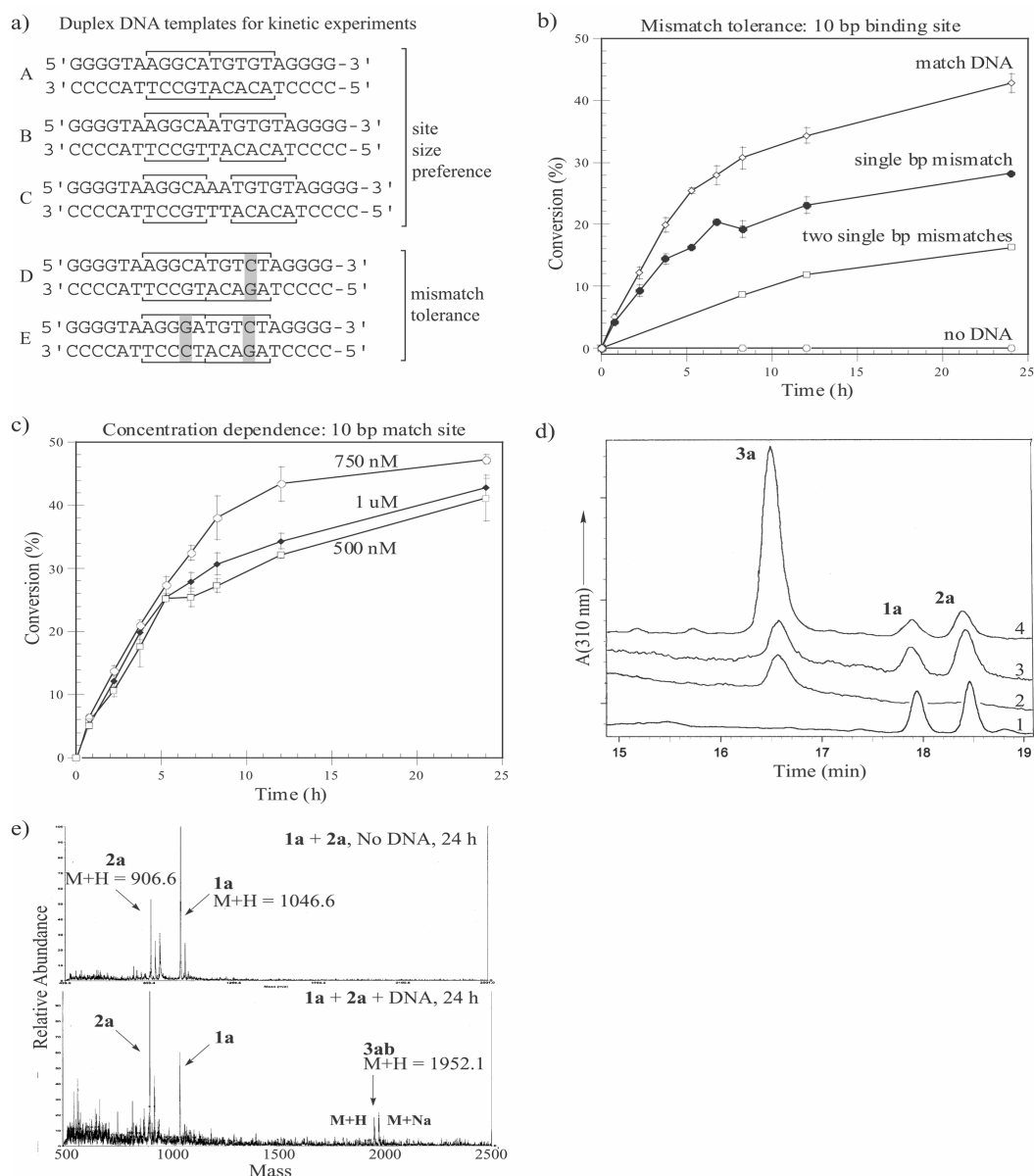
Figure 2.4. a) Sequences of the short DNA duplexes used in examining the kinetics of DNA-templated tandem polyamide formation. Duplexes **A**–**C**: site size preference. Duplexes **D**–**E**: mismatch tolerance at the optimal 10 bp template site length. b) Mismatch tolerance rate data for the formation of tandem **3ab** from **1a** and **2a** at 1 μM concentrations at the 10 bp template site. Open diamond, duplex **A** 5'-AGGCATGTGT-3'; closed circle, duplex **D** 5'-AGGCATGTC̲T-3'; open square, duplex **E** 5'-AGGG̲ATGTC̲T-3'; open circle, no DNA template. Each data point is an average of three kinetics experiments. c) Rate data for formation of tandem **3ab** from **1a** and **2a** on duplex **A** 5'-AGGCATGTGT-3'. Closed diamond, 1 μM; open circle, 750 nM; open square, 500 nM. Each data point is an average of three kinetics experiments. d) HPLC product verification from a single kinetics experiment between **1a** and **2a** on duplex **A** 5'-AGGCATGTGT-3' at 1 μM: Trace 1) **1a** + **2a** on duplex **A**, 0 hours; Trace 2) 1,4-regioisomeric **3a** with duplex **A**; Trace 3) **1a** + **2a** on duplex **A**, 8 hours, 37 °C; Trace 4) **1a** + **2a** on duplex **A**, 8 hours, + authentic sample of 1,4-regioisomeric **3a**. e) MALDI-TOF mass spectrometry product verification: No DNA) Mass spectra of **1a** + **2a** at 1 μM with no DNA, 24 h, 37 °C; + DNA) Mass spectra of **1a** + **2a** at 1 μM with duplex **A** 5'-AGGCATGTGT-3', 24 h, 37°C, arrow highlights mass corresponding to **3ab**.

authentic sample of **3a** (verified as the 1,4-regioisomer) was added. When analyzed by HPLC, no new peaks were observed while the putative product peak grew in absolute magnitude (Figure 2.4d). Purification of reaction mixtures by ZipTip C_{18}-charged pipette tips and subsequent analysis of the 50% CH_3CN eluent shows a peak corresponding to the mass of **3ab** only when the DNA template **A** is present (Figure 2.4e).

Quantitative DNase I footprinting (Tables 2.3 and 2.4). Once it was established that **1a** and **2a** were optimal reaction partners for the templated cycloaddition, the equilibrium association constants and sequence specificity of these hairpin polyamides and the 1,4-regioisomeric tandem product **3a** were analyzed by quantitative DNase I footprinting.[59] The polyamides were characterized on a DNA fragment from pATK1 which contains the sites 5'-AGGCATGTGT-3', 5'-AGGCAATGTGT-3', and 5'-AGGCAAATGTGT-3'. This allowed the comparison of binding affinities for DNA sequences including zero, one, or two intervening base pairs between the hairpin polyamide binding sites. In order to assess the sequence specificity of mismatches on overall tandem binding to the entire 10-bp site, 5'-AGGCATGTGT-3', the polyamides were assayed against a restriction fragment of DNA from pATK2 containing the match site, two single-bp mismatch sites targeted to each half-site (5'-AGGGATGTGT-3' and 5'-AGGCATGTCT-3') and the double-bp mismatch site 5'-AGGGTAGTCT-3' (Figure 2.5).

Six-ring hairpin module **1a** binds to its designed match sites 5'-AGGCA-3' on the restriction fragments from both pATK1 and pATK2 with modest affinity ($K_a = 3.5 \times 10^8$ M^{-1}). On the restriction fragment from pATK2, which contains the mismatch site 5-AGGGA-3', polyamide **1a** shows sequence specificity, favoring the match site over the

mismatch site by >340-fold (Figure 2.6). The affinity compares favorably with that found previously for the parent polyamide ImImPy-γ-ImPyPy-β-Dp ($K_a = 6.2 \times 10^7$ M^{-1}) footprinted at the site 5'-TGGCT-3' on pSES-TL1.[60]
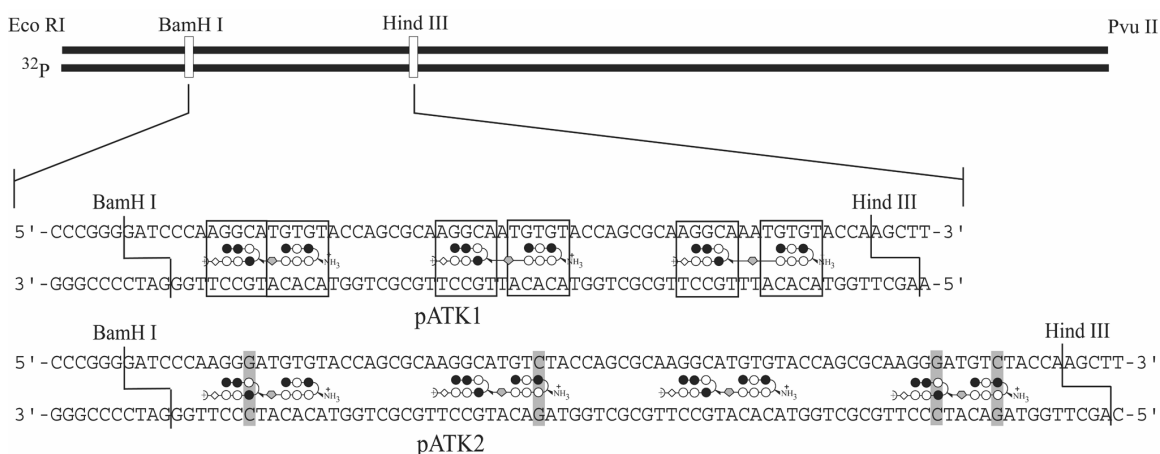
Figure 2.5. Sequences of the *Bam*HI/*Hind*III inserts from the *Eco*RI/*Pvu*II restriction fragments from pATK1 and pATK2 representing the designed sites for quantitative DNase footprinting assays. (See Supplemental Fig. 1 for footprinting gels.) pATK1- sites for the 10-bp (zero intervening bp), 11-bp (one intervening bp), and 12-bp (two intervening bp) match sites. Designed hairpin binding sites are boxed. pATK2- mismatch tolerance at the 10-bp (zero intervening bp) site. Sites for the two single bp mismatches, the match, and the double bp mismatch. Designed mismatch bases are boxed.

Six-ring hairpin module **2a** binds the designed match sites 5'-TGTGT-3' with $K_a = 9.2 \times 10^7$ M^{-1}. On the restriction fragment from pATK2, which contains the mismatch site 5'-TGTCT-3', polyamide **2a** favors its match site by >10-fold (Figure 2.7). Previous studies showed that the parent polyamide, with an additional positive charge, ImPyIm-(R)^{H2N}γ-PyPyPy-β-Dp binds match site 5'-AGTGA-3' on pIK2 with $K_a = 1.2 \times 10^9$ M^{-1}.

The 1,4-triazole regioisomer of the tandem-hairpin **3a** was characterized on the restriction fragment from pATK1 containing the zero intervening bp 5'-AGGCATGTGT-3', one intervening bp 5'-AGGCAATGTGT-3', and two intervening bp 5'-AGGCAAATGTGT-3' match binding sites. Polyamide **3a** binds to both the 10 bp ($K_a = 6.2 \times 10^9$ M^{-1}) and 11
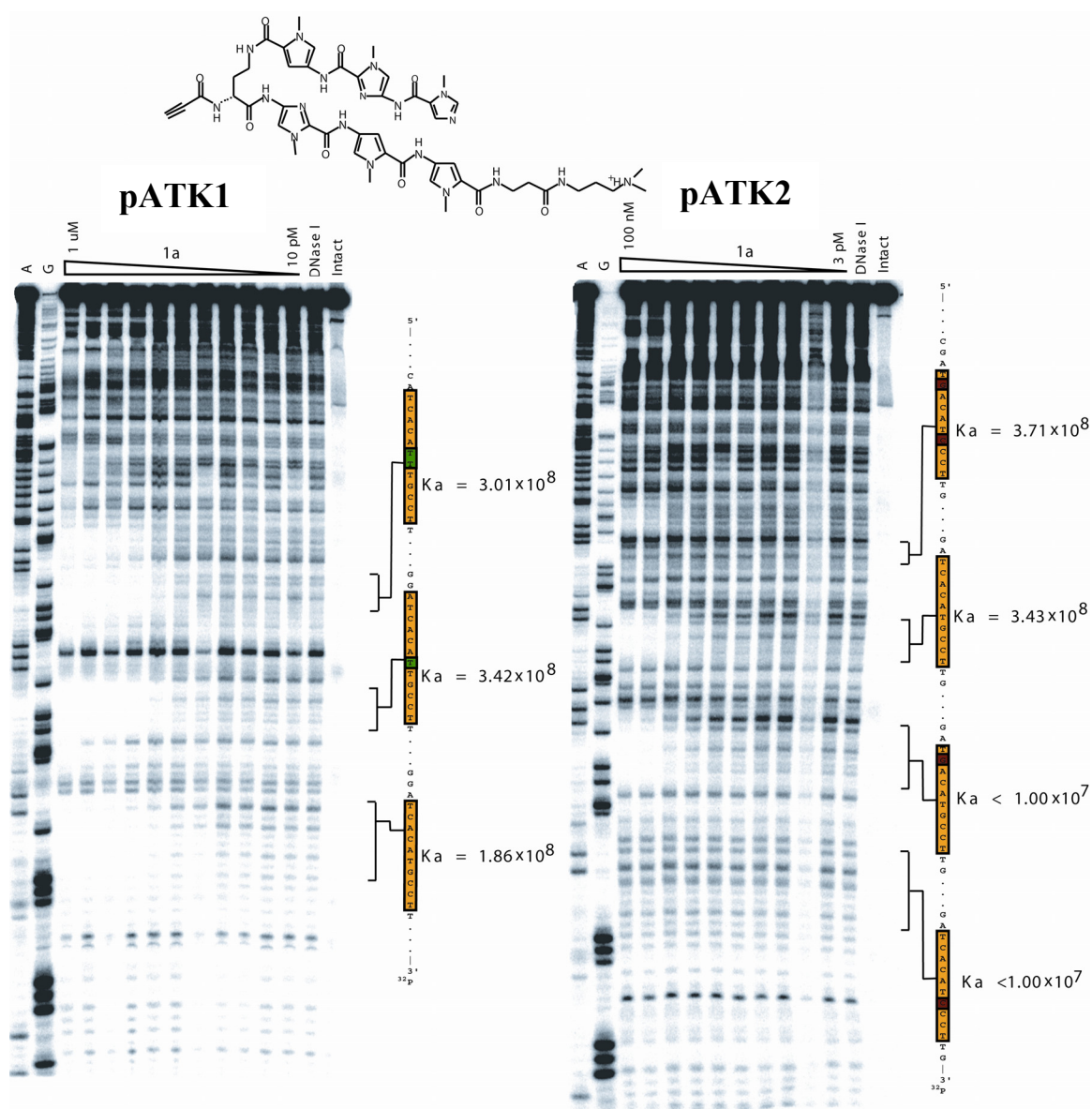
Figure 2.6 Left: Quantitative DNase I footprinting experiments with hairpin **1a** (shown at top) on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK1. From left to right: lane 1, A-specific reaction; lane 2, G-specific reaction; lanes 3–13 1 μM, 300 nM, 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM **1a**; lane 14, DNase I standard; lane 15, intact DNA. Right: Quantitative DNase I footprinting experiments with hairpin **1a** on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK2. From left to right: lane 1, A-specific reaction; lane 2, G-specific reaction; lanes 3–13 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM **1a**; lane 13, DNase I standard; lane 14, intact DNA. All reactions contained 15 Kcpm labeled DNA and were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl2, and 5 mM CaCl2 with an equilibration time of 36 hours.   Designed binding sites where equilibrium association constants were obtained are shown to the right side of the gel.   Equilibrium association constants are listed next to each binding site.
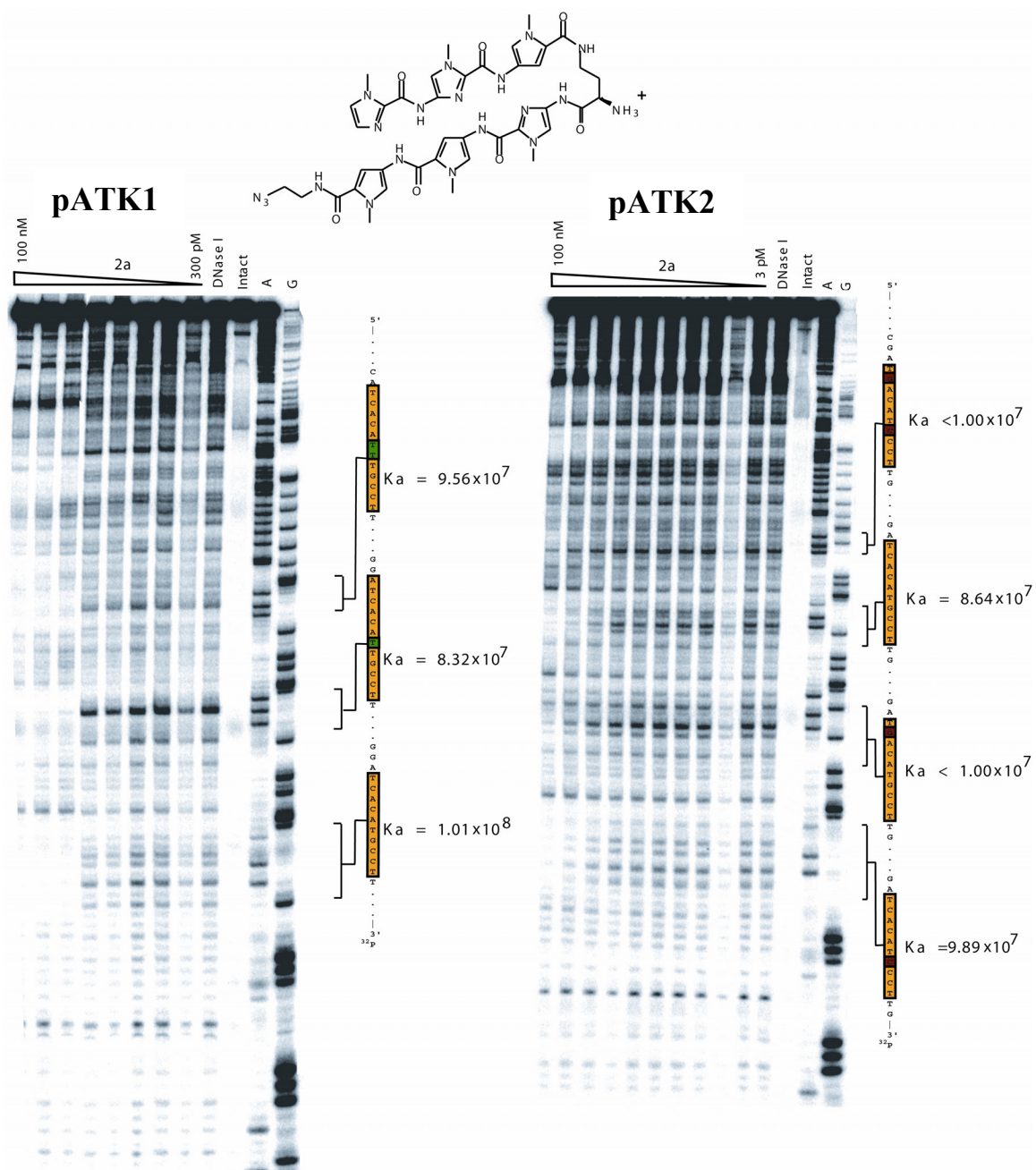
Figure 2.7 Left: Quantitative DNase I footprinting experiments with hairpin **2a** (shown at top) on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK1. From left to right: lanes 1–8 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM **2a**; lane 9, DNase I standard; lane 10, intact DNA; lane 11, A-specific reaction; lane 12, G-specific reaction. Right: Quantitative DNase I footprinting experiments with hairpin **2a** on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK2. From left to right: lanes 1–10 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM **2a**; lane 11, DNase I standard; lane 12, intact DNA; lane 13, A-specific reaction; lane 14, G-specific reaction. All reactions contained 15 Kcpm labeled DNA and were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl2, and 5 mM CaCl2 with an equilibration time of 36 h. Designed binding sites are shown to the right side of the gel. Equilibrium association constants are listed next to each binding site.
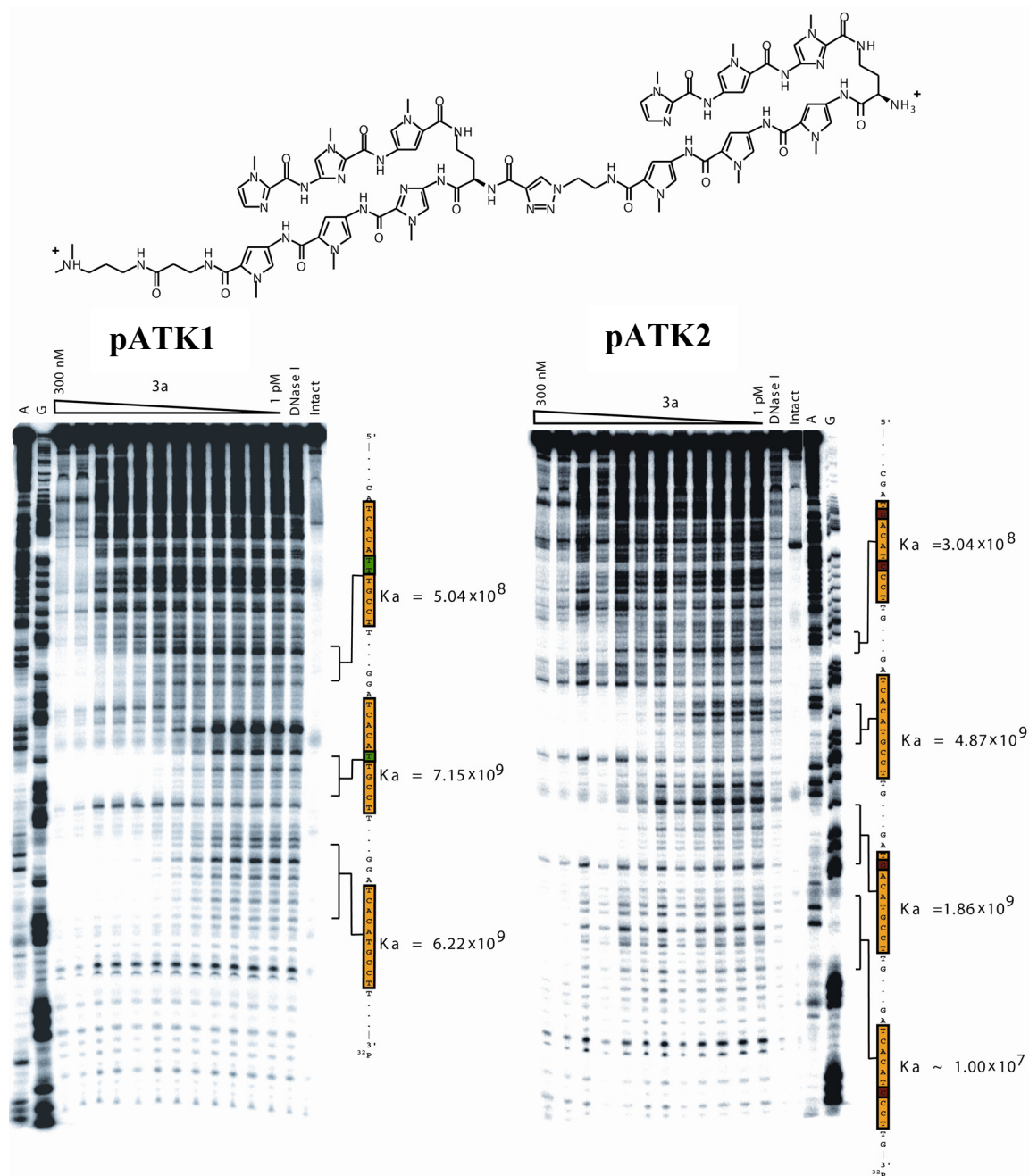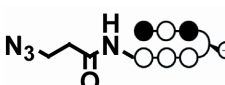
Figure 2.8 Left: Quantitative DNase I footprinting experiments with hairpin dimer **3a** (shown above) on the 3'-^{32}P-labeled DNA fragment derived from plasmid pATK1. Lane 1, A-specific reaction; lane 2, G-specific reaction; lanes 3–14 300 nM, 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM, 1 pM **3a**; lane 15, DNase I standard; lane 16, intact DNA. Right: Quantitative DNase I footprinting experiments with hairpin dimer **3a** on the 3'-32P-labeled DNA fragment derived from plasmid pATK2. Lanes 1–12 300 nM, 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM, 1 pM **3a**; lane 13, Intact DNA; lane 14, DNase I standard; lane 15, A-specific reaction; lane 16, G-specific reaction. All reactions contained 15 Kcpm labeled DNA and were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl2, and 5 mM CaCl2 with an equilibration time of 36 h. Designed binding sites where equilibrium association constants were obtained are shown to the right side of the gel. Equilibrium association constants are listed next to each site.

bp site (K_a = 7.1x10^9 M^{-1}) with high affinity. The size of the footprint indicates that the polyamide is protecting the entire binding site from cleavage by DNase I. At the 12 bp site, the polyamide shows lower affinity (K_a = 5.4x10^8 M^{-1}), while still protecting the entire 12 bp binding site (Figure 2.8).

Table 2.3. Hairpin half-site separation preference. Equilibrium association constants K_a [M^{-1}] for polyamides **1a**, **2a,** and **3a**.[a-c]
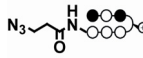
Polyamide	5'-<u>AGGCA</u>TGTGT-3'	5'-<u>AGGCAA</u>TGTGT-3'	5'-<u>AGGCAAA</u>TGTGT-3'
1a	2.8×10^8 ($\pm$0.4)[d]	3.4×10^8 ($\pm$0.4)[d]	3.0×10^8 ($\pm$0.2)[d]
2a	1.0×10^8 ($\pm$0.3)[e]	8.3×10^7 ($\pm$0.2)[e]	9.6×10^7 ($\pm$0.3)[e]
3a	6.2×10^9 ($\pm$0.9)	7.1×10^9 ($\pm$0.4)	5.4×10^8 ($\pm$0.3)

[a] The reported association constants K_a are the average values obtained from three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl$_2$, and 5 mM CaCl$_2$ with an equilibration time of 12 h. [c] Specificities are given in brackets under the K_a values and are calculated as K_a(match)/K_a(mismatch). [d] Affinity constants for **1a** measured on the 5'-aGGCa-3' and 5'-aGGGa-3' sites. [e] Affinity constants for **2a** measured on the 5'-tGTGt-3' and 5'-tGTCt-3' sites.

Mismatch tolerance for the binding of polyamide **3a** was then examined at the 10 bp binding site using the restriction fragment from pATK2 containing the 10 bp (zero intervening bp) match site (5'-AGGCATGTGT-3'), the 10 bp site with a mismatch under **2a** (5'-AGGCATGT<u>C</u>T-3'), the 10 bp site with a mismatch under module **1a** (5'-AGG<u>G</u>ATGTGT-3') and the 10 bp site with a single-bp mismatch under each of the two hairpins (5'-AGG<u>G</u>ATGT<u>C</u>T-3'). Polyamide **3a** again binds its match site with high affinity (K_a = 4.9x10^9 M^{-1}). When a single base pair mismatch is present under the **2a** half site, polyamide **3a** binds with reduced affinity (K_a = 1.9x10^9 M^{-1}). When a single base pair mismatch is present under the **1a** half-site, the binding affinity of **3a** is greatly

reduced ($K_a \leq 1.0 \times 10^7$ M^{-1}). When a single base pair mismatch is present at both hairpin

sites, the equilibrium association constant of **3a** is similarly reduced ($K_a \leq 1.0 \times 10^7$ M^{-1}).

In each case where we measure a binding constant, the entire length of the 10 bp binding

site is protected from cleavage by DNase I.

Table 2.4. Mismatch tolerance at the 10 bp binding site. Equilibrium association constants K_a [M^{-1}] for polyamides **1a**, **2a**, and **3a**.[a-c]

Polyamide	5'-AGGCATGTGT-3'	5'-AGG**G**ATGTGT-3'	5'-AGGCATGT**C**T-3'	5'-AGG**G**ATGT**C**T-3'
1a	$3.7 \times 10^8 (\pm 0.4)^d$	$\leq 1 \times 10^7 (\pm 0.4)^d$ [≥ 37]	$3.4 \times 10^8 (\pm 0.2)^d$	$\leq 1 \times 10^7 (\pm 1.5)^d$ [≥ 37]
2a	$8.6 \times 10^7 (\pm 0.3)^e$	$9.9 \times 10^7 (\pm 0.2)^e$	$\leq 1 \times 10^7 (\pm 0.3)^e$ [≥ 9]	$\leq 1 \times 10^7 (\pm 0.4)^e$ [≥ 9]
	$4.9 \times 10^9 (\pm 0.9)$	$\leq 1 \times 10^7 (\pm 0.4)$ [≥ 490]	$1.6 \times 10^9 (\pm 0.3)$ [3]	$\leq 1 \times 10^7 (\pm 0.4)$ [≥ 490]

[a] The reported association constants K_a are the average values obtained from three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl$_2$, and 5 mM CaCl$_2$ with an equilibration time of 12 h. [c] Specificities are given in brackets under the K_a values and are calculated as K_a(match)/K_a(mismatch). [d] Affinity constants for **1a** measured on the 5'-aGGCa-3' and 5'-aGGGa-3' sites. [e] Affinity constants for **2a** measured on the 5'-tGTGt-3' and 5'-tGTCt-3' sites.

Cell Uptake of Fluorescein-Conjugated Polyamides.

Fluorescein-conjugated analogs of polyamides **1a** and **2a** (**9-17**) were synthesized

according to standard procedures (Figure 2.9).[20, 21] Fluorescein was incorporated either

as the diacetylated 5'-amide, or from fluorescein isothiocyanate (FITC) to yield the

thiourea-linked dye.

Conjugates were screened against a panel of human adherent and nonadherent cell

lines according to literature procedures.[20, 21] The data is summarized in Table 2.5.

Alkyne-functionalized conjugates **9** and **10** were able to localize to the nuclei of most
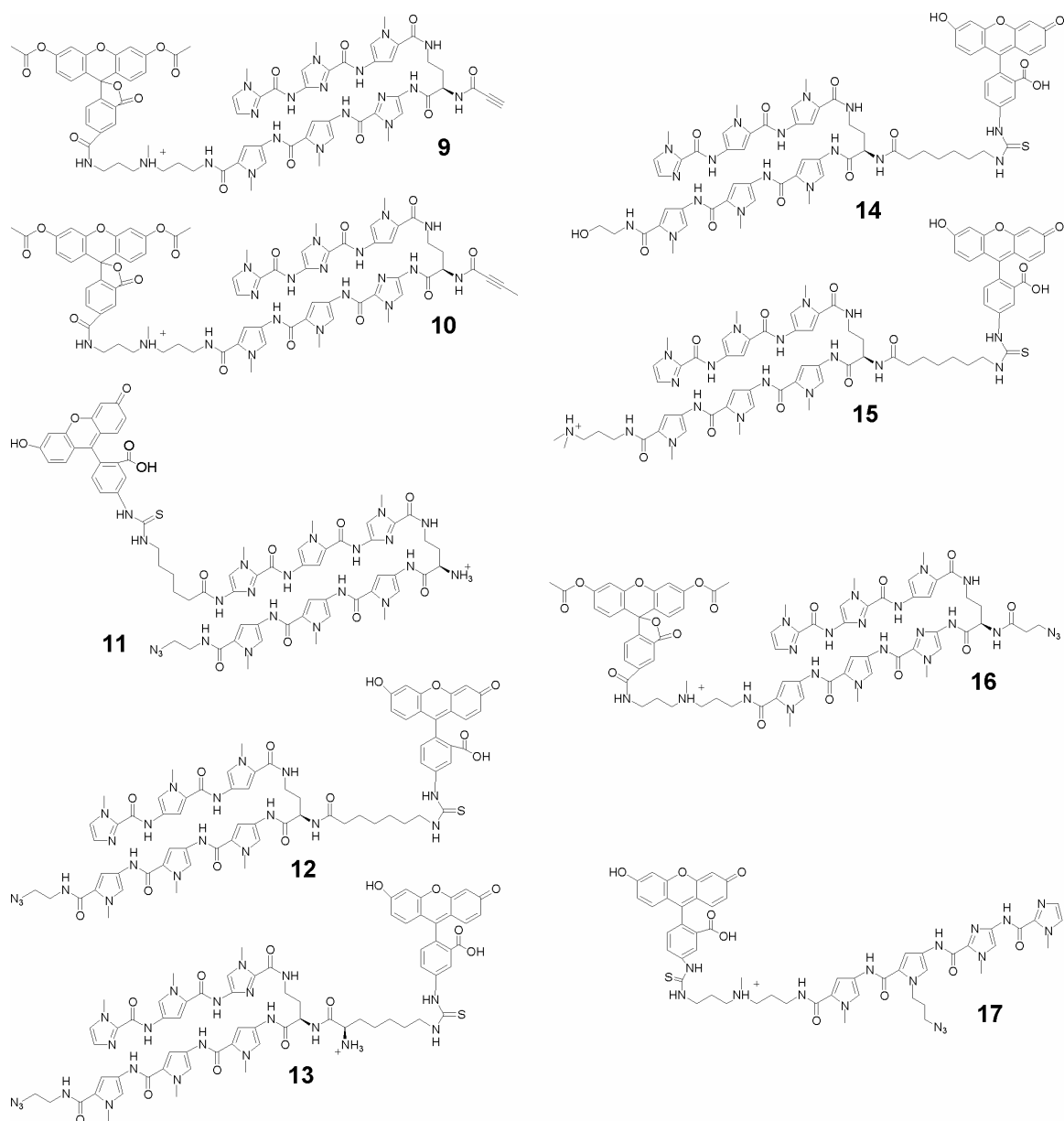
Figure 2.9 Molecular structures of fluorescein dye conjugates synthesized for cell uptake studies.

living cells tested. Conjugate **9**, while showing good nuclear localization, also exhibits

an interesting staining pattern on the surface of the cells. Conjugate **10**, which

is a methyl-protected version of the terminal alkyne **9**, also shows good nuclear

localization without the association with the cell matrix proteins observed with **9**.

		MCF-7	HeLa	PC3	LN-CaP	SK-BR-3	DLD-1	786-O	293	Jurkat	CEM	MEG-01	MEL	NB4
9	Ac₂FAM—(+)—◯◯◯N-propargyl	+	+	+	+	−	+	−−	+	++	++	−	+	++
10	Ac₂FAM—(+)—◯◯◯N-propargyl	++	+	+	+	+	++	+	+	++	+	+	+	+
11	FITC—C₇—structure; N₃—◯◯◯—(+	+	−−	−−	−−	−	−−	−−	−−	−−	−−	−−	−−	−−
12	N₃—◯◯◯—C₇—FITC	+	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
13	N₃—◯◯◯—(+)Lys—FITC	+	−−	−−	−−	−−	−−	−−	−−	−	−	−−	−−	−−
14	HO—◯◯◯—C₇—FITC	+	−−*	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
15	+—◯◯◯—C₇—FITC	−	+*	+	++	+	+	−*	−−	++	++	+	++	+
16	Ac₂FAM—(+)—◯◯◯—N₃	+	+	+	−−	+	+b	−	−−	+	−	−−	−−	−−
17	FITC—(+)—◯◯●● (N₃)	++	−	−	+	+	−−	−−	−−	−	+b	−−	−−	−−

Table 2.5. Cell uptake data for alkyne- and azide-functionalized fluorescein-polyamide conjugates. At left is the schematic representation of each compound. Black and white circles represent imidazole and pyrrole carboxamides, respectively. Data table is filled out with subjective determinations of nuclear localization. Double plus represents a strong nuclear localization; plus represents nuclear accessibility; minus represents weak nuclear localization; double minus represents exclusion from nucleus.

Azide-functionalized conjugates **11**, **12,** and **13** were entirely excluded from the nuclei of most living cells tested, the exception being MCF-7 (kidney cancer) cells, which allow modest uptake (Figure 2.10).

In order to determine whether the azide moiety is responsible for the lack of nuclear localization, controls **14** and **15**, which differ from **12** only in the tail functionality, were assayed against the panel of cells. Terminal alcohol conjugate **14** exhibits a pattern of localization similar to azide **12**, while Dp-functionalized **15** shows strong nuclear localization in many of the cell lines tested.

In order to verify that molecular geometry was not the cause of conjugates **11**–**13** being excluded from the nuclei, conjugate **16** was synthesized. This polyamide possesses

9 - jurkat cells

11 - jurkat cells

9 - MCF-7 cells

11 - MCF-7 cells
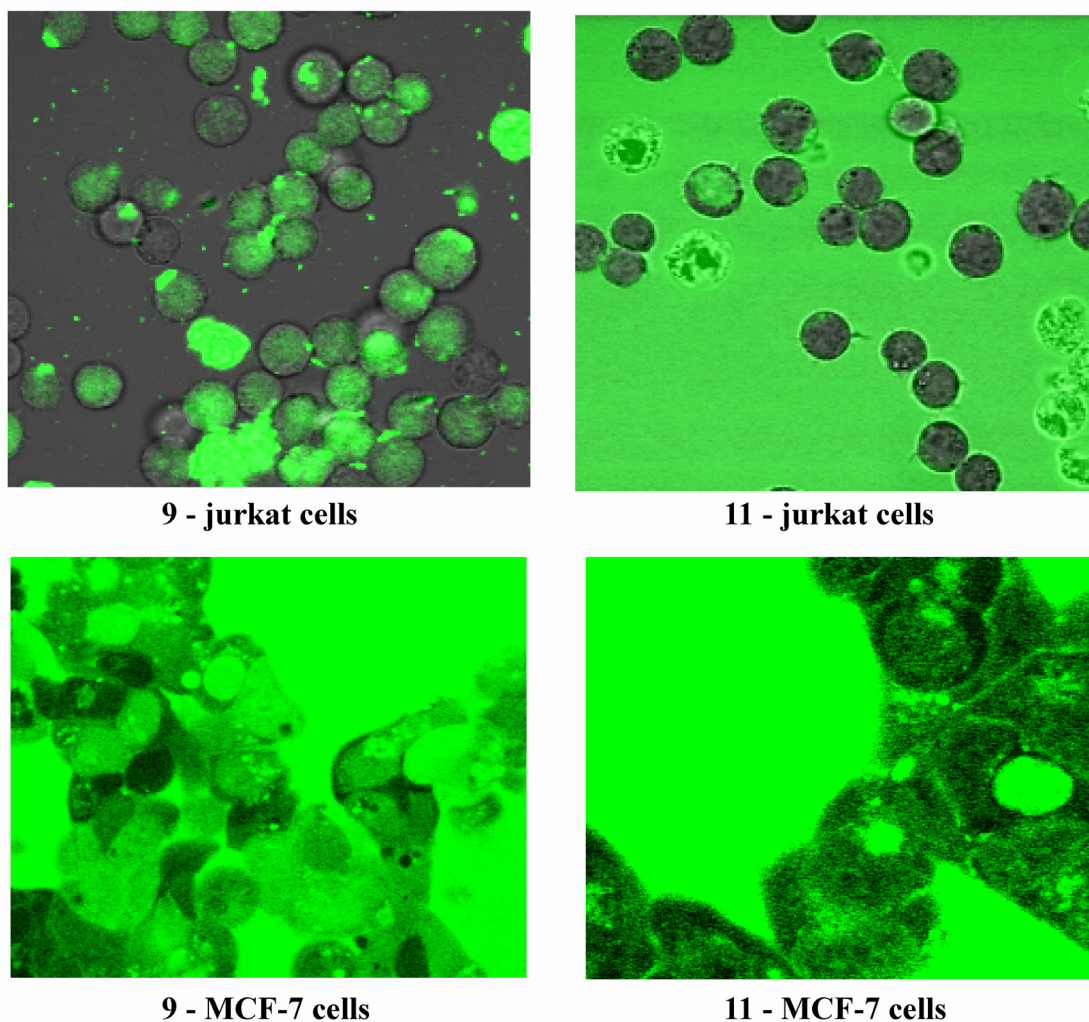
Figure 2.10. Fluorescent images taken with a confocal microscope. At top are shown images with compound **9** and **11** in jurkat cells. Compound **9** shows clear nuclear localization and is rated a "++." Compound **11** is completely excluded from the nuclei, and is rated a "- -." At bottom is shown images with compounds **9** and **11** in mcf-7 cells. Both compounds show some nuclear localization while the majority of the fluorescence is outside the cells and are rated "+." Cells with homogeneously bright areas are dead.

identical geometry to conjugates **9** and **10**, with the dye conjugated to the C-terminal tail and the azide conjugated to the turn amine. This polyamide was able to access the nuclei in several cell lines.

As a final control, the short, 4-ring polyamide-azide-dye conjugate **17** was synthesized. Interestingly, this smaller molecule was still only able to access the nuclei of a few cells.

Discussion

This work represents an exploratory effort towards using the minor groove of double-helical DNA to template a chemical reaction wherein the polyamide product encodes the base sequence content of the DNA template. The DNA-templated tandem polyamide formation occurs through the Hüisgen 1,3-dipolar cycloaddition reaction in aqueous media at 37 °C (pH = 7.0). The minor groove of DNA appears to impart some additional steric constraint upon the 1,3-dipolar cycloaddition. In each pairing, the product ratio of regioisomers is increased in the templated versus the non-templated reaction. While non-templated pairings of the activated alkyne **1a** with either **2a** or **2b** resulted in a 20:1 ratio of regioisomers, these reactions, when templated by DNA, produce the major thermal product exclusively. Non-templated couplings of unactivated alkyne **1b** resulted in formation of both regioisomers in equal amounts. However, when templated on DNA, the ratio was increased to the point where only a single isomer was produced (**1b** with **2a**, duplex **A**; and **1b** with **2b**, duplex **B**), or to a ratio of 3:1 (**1b** with **2b**, duplex **A**). These results can be rationalized by the fact that the 1,4-regioisomer is formed by an antiparallel approach of the two reactants while the 1,5-regioisomer is formed from a parallel approach of the two reactants. When the reactive partners are sequestered in the minor groove of DNA, the linkers are forced to span the space between the two hairpin binding sites and approach each other in an antiparallel fashion, thereby

favoring the pathway leading to the 1,4-regioisomer. The fact that any of the minor isomer is produced by the longest linkers on the shortest template suggests that the space between the hairpin binding sites is short enough for the long linkers to approach each other in a parallel fashion while still being close enough to react.

Both the activated alkyne (**1a**) and the alkyl alkyne (**1b**) were unreactive toward either azide (**2a–b**) in the absence of DNA template up to 1 μM concentrations. Hairpins **1a** and **2a** at similar concentrations form tandem **3a** on a template **A** with a rate increase of greater than 10,000-fold over the non-templated version. Hairpin **1a** also forms tandem- dimer **4ab** on the 10 bp duplex **A** (zero intervening base pairs) when paired with the longer, more flexible azide **2b**. However, the rate of tandem formation was slower than the rate of formation from **1a** and **2a**. This decreased rate is perhaps due to the additional flexibility in the linker, which allows the reactants more freedom to adopt non-productive conformations. Hairpin **1b**, which exhibits a longer alkyne linkage (2 additional methylene units), was also tested with the azido-functionalized hairpins **2a** and **2b**. Both pairings resulted in tandem dimer formation on the 10 bp site. The pairing of **1b** and **2b** also showed product formation on the longer 11 bp site, with the 10 bp site preferred by a ratio of 6:1. However, the rate of product formation by these pairings was 800 times slower than the rate of formation between **1a** and **2a** on the 10 bp binding site. The decreased rate may be due to the inherent differences in reactivity between **1a** and **1b**. Thus, while DNA template **A** increases the rate of cycloadditions between **1b** and either azide more than two orders of magnitude, the reaction still takes weeks to proceed to moderate yields. The rate of DNA-templated cycloaddition with the activated alkynyl

amidate **1a** is increased more than 10,000-fold by template **A**, which causes the reaction to proceed in hours, a timescale that is relevant to biological applications.

The double-helical DNA-templated cycloaddition also shows a dependence on the flexibility of the linkers between the chemical reactants and the DNA-binding domains. The longer, more flexible linkers (**1b** + **2b**) allow the cycloaddition to proceed at both the 10 bp and 11 bp template sites. However, the more restricted linker (**1a** + **1b**) limits the intervening distance over which cycloaddition may occur to a single bp. Both reactivity and orientation of the azide and alkyne moieties appears critical to optimal reactivity.

The cycloaddition reaction is also shown to be dependent not only upon the spacing between the hairpins, but also upon the sequence composition of the DNA under the template recognition elements of the **1a** and **2a** optimal pair. When a single base-pair mismatch is present under the azide-functionalized hairpin polyamide **2a** at the 10 bp binding site, the rate of the tandem-forming cycloaddition is nearly halved. Additionally, when a single base-pair mismatch is introduced under each of the hairpin modules at the 10 bp binding site, the cycloaddition reaction is reduced 30-fold.

When the concentration of reacting species is varied from 1 μM to 500 nM, the initial rates of cycloadduct **3a** formation do not significantly change, which implies that pseudozero-order kinetics apply to this reaction. While the rate equations that describe this process fully are likely complex (and not the focus of this paper), polyamide association to DNA is known to be near diffusion limited.[61] Once the polyamides occupy both half-sites on DNA, the cycloaddition is an *intramolecular* reaction that competes with complex dissociation. We believe that at concentration regimes near the IC_{50} values for **1a** and **2a** (~10 nM), the kinetics would reflect the DNA bound intramolecular

cycloaddition reaction. It is clear from the DNase I footprint titration experiments that at concentrations between 1 μM and 500 nM, hairpins **1a** and **2a** cannot fully distinguish their match and mismatch sites, likely occupying both sites. Thus, tandem **3a** formation observed on duplex **D** and **E** (10 bp site, 1 or 2 mismatch bases, respectively) is most likely a consequence of complex formation despite the presence of mismatch DNA. Furthermore, the fact that the reactions proceed to less than 50% yield is most likely a consequence of the ability of polyamides **1a** and **2a** to occupy mismatch DNA sites and thus bind in non-product-forming orientations. The kinetic results on the templated formation of **3a** on duplex **E** (10 bp site, two single bp mismatches) further support this hypothesis. At concentrations of 1 μM, **3a** is formed in 16% yield, while at 750 nM and 500 nM concentrations **3a** is formed in only 8% yield. This suggests a threshold exists between 1 μM and 750 nM that defines the ability of these polyamides to recognize their match site in preference to the double base pair mismatch site. It follows that at some lower concentration, there is another threshold that allows the templated cycloaddition to discriminate its match site from a single-bp mismatch. Thus, at concentrations at which these polyamides can fully distinguish their match site, the ratio of product formation on match versus mismatch DNA as well as the overall tandem yield should increase.

Once establishing that the pair of hairpins **1a** and **2a** showed the most favorable template-directed cycloaddition with respect to rate and specificity, the binding properties of the hairpin starting materials and tandem product **3a** were analyzed by quantitative DNase I footprinting. The tandem dimer formed between **1a** and **2a** shows a >12-fold increase in binding affinity over either of the two hairpin starting materials. At the 10 bp site, tandem **3a** exhibits only a modest 3-fold specificity for its match site over a single

base pair mismatch under the azide-functionalized **2a**, but shows good specificity over a single base pair mismatch under the alkyne-functionalized **1a** and over a double base pair mismatch (>400 fold). Dimer **3a** targets both the 10 and 11 bp sequences of DNA with high affinity and specificity. However, because the product is formed solely at the 10 bp site, and subsequently exhibits high affinity for that site, its rate of dissociation from that match site should be very slow, rendering it specific for the 10 bp site.

Finally, because our eventual goal is to form these tandem structures inside living cells, nuclear localization of alkyne and azide-functionalized polyamide-fluorescein conjugates was assayed. The alkyne-functionalized polyamides **9** and **10** were able to strongly localize to the nuclei of most cells tested. Interestingly, the terminal alkyne showed some staining along the cellular membrane. Perhaps the alkyne moiety is undergoing some Michael-type reaction with cell matrix proteins, thus causing the unusual staining pattern. Protection of the terminal alkyne with a methyl group (compound **10**) led to a molecule with good nuclear localization and decreased membrane staining.

Compounds functionalized with an azide at the tail were unable to access the nuclei of most cells tested. That control compound **15**, with the Dp tail in place of the ethyl azide, shows strong nuclear localization in many cell lines indicates that the azide functionality of compounds **11–13** may be a negative determinant for polyamide uptake. That compound **16** localized to the nuclei of several cell lines indicates that the placement of the azide group is important for cell uptake, and that polyamides with the azide moiety conjugated to the γ-aminobutyric acid turn residue possess better uptake properties than their C-terminal azide counterparts.

Conclusion.

Double-helical DNA is capable of templating the site-specific formation of tandem-hairpin dimers. At μM concentrations, the reaction between hairpin polyamides **1a** and **2a** in the DNA minor groove exhibit a 10,000-fold rate enhancement. The templated 1,3-dipolar cycloaddition reaction is sensitive to separation distance between adjacent DNA binding sites. The Watson-Crick sequence information in the DNA helix is encoded by the sequence composition in the pyrrole-imidazole polyamide product. In this supramolecular system, the input molecule is DNA and the output molecule is an organic product with "improved function" with regard to DNA recognition properties (i.e., increased binding-site size and higher affinity). By extension, regarding the design of chemical systems that ligate nonenzymatically in the minor groove of DNA in a sequence dependent fashion,[46] other reactive functional pairs can be considered— such as the Diels-Alder reaction, Staudinger ligation, and S_N2-type reactions— that fit the criteria of thermal reactivity at 37 °C in water with orthogonality towards chemical moieties found inside living systems.

Regarding self-assembly in live cells, we anticipate at least two technical hurdles. Cell uptake remains a large hurdle.[20, 21] Because the azide- functionalized polyamides do not exhibit favorable uptake properties, other ligation reactions, such as those listed above, that utilize polyamides with different functionality, may have to be used. Alternatively, because the turn-linked azide compound did show improved cell uptake, perhaps azide and alkyne pairs can be used for biological templating reactions if

molecules are designed to incorporate the azide only at this position. Secondly, when the starting materials are presented with a gigabase of DNA, will the polyamides be able to avoid the large number of single- and double-bp mismatch sites in order to equilibrate to their unique contiguous match sites? Perhaps we may need to start by targeting repeat regions of DNA such as the centromeres or telomeres that are known to aggregate polyamides.[62] In may be necessary to move to longer, more specific polyamides as starting materials. Finally, the DNA target is condensed one-millionfold on chromatin, and ligation may be better attempted across adjacent minor grooves between two aligned superhelical gyres ("supergrooves" as reaction platforms).[63, 64]

Experimental

Materials

Boc-β-alanine-(4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-ala-PAM resin) was purchased from Peptides International. Oxime resin was purchased from Nova Biochem. Oligonucleotides for kinetics and footprinting were purchased from the Biopolymer Synthesis and Analysis Facility at the California Institute of Technology. Propiolic acid, sodium azide, and methanesulfonyl chloride were purchased from Aldrich. All other synthetic and footprinting reagents were as previously described.

^{1}H NMR spectra were recorded on a Varian Mercury 300 instrument. UV spectra were recorded on a Beckman Coulter DU 7400 diode array spectrophotometer. Autoradiography was performed with a Molecular Dynamics Typhoon PhosphorImager. MALDI mass spectra were obtained on a Voyager De PRO time-of-flight mass

spectrometer (Applied BioSystems) operated at an accelerating voltage of +20 dV. The samples were dissolved in 50% CH_3CN: 0.1% $TFA-H_2O$ and applied to the target in a α-cyanohydroxycinnamic matrix. The mass spectrometer was calibrated with a calibration mixture provided by the instrument manufacturer. DNA sequencing was performed at the Sequence/Structure Analysis Facility (SAF) at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed-phase column in 0.1% (w/v) $TFA-H_2O$ with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% CH_3CN/min. Preparatory HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100 μm C_{18} column, 0.1% (w/v) $TFA- H_2O$, 0.4% CH_3CN/min. Water was obtained from a Millipore MilliQ water purification system and all buffers were 0.2 μm filtered. All reagent-grade chemicals were used without further purification unless otherwise stated.

ImImPy-(R)^{H_2N}γ-ImPyPy-β-Dp (4) Synthesized on solid support according to literature procedures.[53]   UV (H_2O) λ_{max} 310 nm (51540).   MALDI-TOF-MS calcd. for $C_{45}H_{60}N_{19}O_8$ (M + H): 994.5. Found 994.6.

ImImPy-(R)[HC≡COC]^{HN}γ-ImPyPy-β-Dp (1a)   Synthesized from **4** according to modified literature procedures.[53] **4** (30 μmol, 30 mg) was dissolved in 1 mL DMF and 1.5 mL acetonitrile and cooled to 0 °C. Propiolic acid (30 μmol, 2.1 mg) was added, followed by DCC (30 μmol, 6.2 mg), and the solution stirred for 15 min at 0 °C then 1 h at rt. Pure product was obtained after purification by reversed-phase HPLC, followed by

lyophilization of appropriate fractions as a white powder (7.8 mg, 7.5 μmol, 25% recovery); UV (H₂O) λ_{max} 310 nm (51540). ¹H NMR (DMSO-d_6) δ 10.35 (s, 1H), 10.31 (s, 1H), 10.13 (s, 1H), 9.90 (s, 1H), 9.71 (s, 1H), 9.22 (bs, 1H), 9.13 (d, 1H, J=4.8 Hz), 8.03 (t, 2H, J=5.7 Hz), 8.02 (m, 2H), 7.55 (s, 1H), 7.45 (s, 1H), 7.44 (s, 1H), 7.43 (s, 1H), 7.24 (d, 1H, J=1.8 Hz), 7.21 (d, 1H, J=1.8 Hz), 7.15 (d, 1H, J=1.5 Hz), 7.13 (d, 1H, J=2.1 Hz), 7.06 (s, 1H), 6.99 (d, 1H, J=1.5 Hz), 6.86 (d, 1H, J=2.1 Hz), 4.51 (q, 1H, J=6.3 Hz), 4.23 (s, 1H), 3.99 (s, 6H), 3.94 (s, 3H), 3.82 (s, 3H), 3.78 (s, 6H), 3.36 (q, 2H, J=5.7 Hz), 3.22 (m, 2H), 3.09 (q, 2H, J=6.0 Hz), 2.98 (m, 2H), 2.72 (d, 6H, J=4.8 Hz), 2.33 (t, 2H, J=7.2 Hz), 1.95 (m, 2H), 1.72 (m, 2H), MALDI-TOF-MS calcd. for $C_{48}H_{60}N_{19}O_9$ (M + H): 1046.5. Found 1046.6.

ImImPy-(R)[HC≡C(CH₂)₂OC]^{HN}γ-ImPyPy-β-Dp (1b)   Synthesized from **4** (3 μmol, 3.1 mg) using a similar procedure to the synthesis of **1a**, substituting 1-pentynoic acid (10 μmol, 1.0 mg) for propiolic acid. Product was obtained as a white powder (2.2 mg, 2.1 μmol, 70% recovery); UV (H₂O) λ_{max} 310 nm (51540).   MALDI-TOF-MS calcd. for $C_{50}H_{64}N_{19}O_9$ (M + H): 1074.5. Found 1074.4.

ImPyIm-(R)^{BocHN}γ-PyPyPy-(CH₂)₂-OH (5a) Product obtained as a white powder upon aminolysis from oxime resin using ethanolamine (neat) (35 mg, 40 μmol, 8% recovery); UV (H₂O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. for $C_{45}H_{57}N_{16}O_{10}$ (M + H): 981.4. Found 981.2

ImPyIm-(R)^{BocHN}γ-PyPyPy-(CH$_2$)$_2$-OMs (6a) Synthesized from **5a** according to modified literature procedures. **5a** (40 μmol, 35 mg) was dissolved in 750 μL CH$_2$Cl$_2$ and cooled to 0 °C. DIEA (0.3 mmol, 38 mg) was added and the solution stirred for 15 min. Methanesulfonyl chloride (90 μmol, 15 mg) was added and the reaction stirred for 10 min at 0 °C then 1 h at room temperature. The solvent was removed by evaporation and the compound purified by reversed-phase HPLC. Product was obtained as a white powder upon lyophilization of the appropriate fractions (30.7 mg, 32 μmol, 80%); UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. for C$_{46}$H$_{59}$N$_{16}$O$_{12}$S (M + H): 1059.4. Found 1059.6.

ImPyIm-(R)^{H_2N}γ-PyPyPy-(CH$_2$)$_2$-N$_3$ (2a) Compound **6a** (32 μmol, 30.7 mg) was dissolved in 750 μL DMF. Sodium azide (1 mmol, 65 mg) was added and the solution stirred at 65 °C for 12 h.[56] The solvent was evaporated and the residue taken up in 1 mL 50% TFA-CH$_2$Cl$_2$ and allowed to stand for 10 min. Pure product was obtained by reversed-phase HPLC purification followed by lyophilization of the appropriate fractions (9.06 mg, 10 μmol, 31% over 2 steps); UV (H$_2$O) λ_{max} 310 nm (51540). ^{1}H NMR (DMSO-d_6) δ 10.50 (s, 1H), 10.41 (s, 1H), 10.11 (s, 1H), 9.97 (s, 1H), 9.92 (s, 1H), 8.35 (t, 2H, J=5.7 Hz), 8.25 (m, 4H), 7.50 (s, 1H), 7.39 (s, 2H), 7.24 (s, 1H), 7.22 (s, 1H), 7.19 (s, 1H), 7.14 (s, 1H), 7.03 (s, 2H), 6.92 (s, 1H), 6.88 (s, 1H), 3.97 (s, 3H), 3.92 (s, 3H), 3.82 (s, 9H), 3.78 (s, 3H), 3.39 (m, 3H), 2.99 (m, 2H), 1.99 (m, 2H), 1.21 (t, 2H, J=6.0 Hz). MALDI-TOF-MS calcd. for C$_{40}$H$_{48}$N$_{19}$O$_7$ (M + H): 906.4. Found 906.6.

ImPyIm-(R)^{H_2N}γ-PyPyPy-(CH$_2$)$_3$-N$_3$ (2b) Compound **2b** was prepared according to the

procedure for **2a**, substituting 3-aminoethanol for ethanolamine for the nucleophilic resin

cleavage. Product was obtained in 2.3% overall yield from resin bound starting material;

UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. for C$_{41}$H$_{50}$N$_{19}$O$_7$ (M + H):

920.2. Found 920.4.

ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_2$-Tr-(OC)]^{HN}γ-ImPyPy-β-Dp   **(3a)**

Compounds **1a** (5 µmol, 5.23 mg) and **2a** (5 µmol, 4.5 mg) were dissolved in 1 mL 20%

(v/v) acetonitrile in water. The solution was lyophilized and the powder formed into a

tight pellet. The pellet was heated under argon for 6 days at 55 °C. Product was obtained

as a white powder after reversed-phase HPLC purification and lyophilization of the

appropriate fractions (800 nmol, 1.56 mg, 16%). UV (H$_2$O) λ_{max} 310 nm (103080). ^{1}H

NMR (DMSO-d_6) δ 10.51 (s, 1H), 10.41 (s, 1H), 10.38 (s, 1H), 10.29 (s, 1H), 10.13 (s,

1H), 10.11 (s, 1H), 9.97 (s, 1H), 9.91 (s, 2H), 9.73 (s, 1H), 9.33 (bs, 1H), 8.63 (s, 1H),

8.56 (d, 1H, J=7.5 Hz), 8.35 (t, 2H, J=5.5 Hz), 8.33 (d, 2H, J=4.5 Hz), 8.29 (m, 2H),

8.19 (t, 1H, J=5.5 Hz), 8.06 (t, 2H, J=6.0 Hz), 8.03 (t, 2H, J=5.5 Hz), 7.56 (s, 1H), 7.51

(s, 1H), 7.47 (s, 1H), 7.46 (s, 1H), 7.40 (s, 1H), 7.39 (s, 1H), 7.24 (s, 1H), 7.22 (s, 2H),

7.17 (s, 1H), 7.15 (s, 2H), 7.14 (s, 1H), 7.07 (s, 1H), 7.04 (s, 2H), 6.98 (s, 1H), 6.29 (s,

1H), 6.88 (s, 1H), 6.87 (s, 1H), 4.73 (m, 4H), 4.59 (m, 5H), 4.00 (s, 3H), 3.99 (s, 3H),

3.98 (s, 3H), 3.95 (s, 3H), 3.93 (s, 3H), 3.85 (s, 6H), 3.83 (s, 3H), 3.80 (s, 3H), 3.78 (s,

3H), 3.77 (s, 3H), 3.65 (m, 2H), 3.37 (m, 2H), 3.23 (m, 2H), 3.11 (q, 2H, J=6.0 Hz), 3.01

(m, 2H), 2.75 (s, 3H), 2.74 (s, 3H), 2.35 (t, 2H, J=6.5 Hz), 2.09 (m, 2H), 2.00 (m, 6H),

1.72 (qu, 2H, J=8.0 Hz), 1.23 (m, 6H). MALDI-TOF-MS calcd. for $C_{88}H_{107}N_{38}O_{16}$ (M + H): 1952.8. Found 1952.1.

ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_3$-Tr-(OC)]^{HN}γ-ImPyPy-β-Dp (4ab)

Prepared from **1a** and **2b** according to the procedure for **3a** (15% yield); UV (H$_2$O) λ$_{max}$ 310 nm (103080). MALDI-TOF-MS calcd. for $C_{89}H_{109}N_{38}O_{16}$ (M + H): 1966.9. Found 1966.7.

ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_2$-Tr-(CH$_2$)$_2$(OC)]^{HN}γ-ImPyPy-β-Dp (5ab)

Prepared from **1b** and **2a** according to the procedure for **3a** (29% yield); UV (H$_2$O) λ$_{max}$ 310 nm (103080). MALDI-TOF-MS calcd. for $C_{90}H_{111}N_{38}O_{16}$ (M + H): 1980.9. Found 1980.5.

ImImPy-(R)-[ImPyIm-(R)^{H_2N}γ–PyPyPy-(CH$_2$)$_3$-Tr-(CH$_2$)$_2$(OC)]^{HN}γ-ImPyPy-β-Dp (6ab)

Prepared from **1b** and **2b** according to the procedure for **3a** (25% yield); UV (H$_2$O) λ$_{max}$ 310 nm (103080). MALDI-TOF-MS calcd. for $C_{91}H_{113}N_{38}O_{16}$ (M + H): 1994.9. Found 1994.8.

Polyamides 9 and 10.

Resin-bound, Fmoc-protected precursor polyamide was treated with 20% piperidine in DMF for 20 min. The resin was washed and then treated with 5 mole

equivalents of t-butoxycarbonyl (BOC) anhydride in 4:1 DMF:DIEA for 2 h. BOC-protected polyamide was liberated from resin with neat 3,3'-N-methyaminodipropylamine. Cleaved product was purified by reversed phase HPLC

(25 % yield). Product was then reacted with 5'fluorescein carboxylic acid diacetate (1.2 mole equivalents), PyBOP (1.5 equiv.), in 4:1 DMF:DIEA. When reaction was complete (by analytical HPLC), the solvent was removed *in vacuo*, and the residue taken up in 50% TFA:DCM for 20 min. The reaction was dried, and the purified by reversed-phase HPLC (70% yield). Final alkyne functionalization was performed as for compounds **1a** and **b** (5% overall yield). UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): **9** 1461.56, **10** 1475.50. Found **9** 1461.8, **10** 1475.9.

Polyamides 11 and 12.

Resin-Bound, Fmoc-protected polyamide precursors were treated with 20% piperidine in DMF for 20 min. 6-amino hexanoic acid (**11**) or lysine (**12**) were then conjugated to the liberated amine using standard amide bond–forming conditions. The polamides were then liberated from resin in neat ethanolamine, and purified by reversed-phase HPLC (20% yield). The terminal hydroxyl groups were transformed into azides according to the procedure for **2a**. The compounds were then coupled to FITC for an overall yield of 2%. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): **11** 1421.56, **12** 1437.54. Found **9** 1421.7, **10** 1437.7.

Polyamide 13.

Boc-Im-OH was coupled in place of the terminal Im-OH cap during standard solid phase synthesis of the precursor polyamide. 6-amino hexanoic acid was then coupled. Fluorescein diacetate was then coupled onto the hexanoic acid. The polyamide-dye conjugate was then liberated from resin using ethanolamine, and the azide installed according to the procedure for **2a**. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1424.49. Found 1424.6.

Polyamides 14 and 15.

Made according to the procedure for **2a**, using ethanolamine (**14**) or Dp (**15**) to cleave the polyamide from resin. FITC was used to conjugate the dye to the turn (12% overall yield). UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): **14** 1396.49, **15** 1438.59. Found **14** 1396.8, **15** 1438.9.

Polyamide 16.

On resin, 3-azido propionic acid was conjugated to the turn amine. Polyamide was then liberated from resin using 3,3'-N-methylaminodipropylamine. FITC was then used to conjugate the dye to the tail (18% overall yield). UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1506.52. Found 1506.7.

Polyamide 17.

Core polyamide was built using N-propanol, Boc-Py-OH. Polyamide was cleaved from resin using mono-BOC-protected 3,3'-N-methylaminodipropylamine. The

hydroxyl was transformed into an azide according to the procedure for **2a**. The BOC group was removed in TFA:DCM, and FITC used to couple the dye onto the C-terminal amine (10% overall yield). UV (H_2O) λ_{max} 310 nm (34360). MALDI-TOF-MS calcd. (M + H): 1080.16. Found 1080.9.

Preparation of duplex DNA for kinetic experiments. Duplex DNA was prepared by incubating equal amounts of complementary sets of synthetic oligonucleotides at 90 °C for 10 min, then slowly allowing them to cool to rt. Resulting duplex DNA was quantified by UV by the relationship 1 OD_{260} unit = 50 µg/mL duplex DNA. Duplex DNA was stored at -20 °C in water.

Cycloaddition reactions. All kinetic reactions were performed in 1.7 mL presiliconized microcenterfuge tubes obtained from VWR International. Total reaction volumes were 1.2 mL aqueous solutions of equimolar concentrations of each hairpin polyamide and DNA (2 mM Tris-HCl, 2 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.0, 37 °C). Reactions were monitored by HPLC by direct injection of reaction samples onto a RAINEN C_{18}, Microsorb MV, 5 µm, 300 x 4.6 mm reversed-phase column in 0.1% (w/v) TFA-H_2O with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 0.5% CH_3CN/min. Peaks were quantified using the Beckman Coulter GOLD software package. Verification of product was determined by MALDI TOF-MS of ~40 µL samples of each reaction concentrated on a ZipTip 2 mg C_{18} pipette tip eluted with 75% (v/v) acetonitrile in 0.1% (w/v) TFA.

Construction of plasmid DNA. Plasmids pATK1 and pATK2 were prepared by hybridization of complementary sets of synthetic oligonucleotides. The hybridized inserts were individually ligated into *Bam*HI/*Hind*III linearized pUC19 using T4 DNA ligase. *E. coli* JM109 high efficiency competent cells were then transformed with the ligated plasmid. Plasmid DNA from ampicilin-resistant white colonies was isolated using a Qiagen Wizard MidiPrep kit. The presence of the desired insert was determined by dideoxy sequencing. Concentration of prepared plasmid was determined by UV by the relationship 1 OD_{260} unit = 50 µg/mL duplex DNA.

Preparation of ^{32}P-end-labeled restriction fragments. Plasmids pATK1 and pATK2 were linearized with *Eco*RI and *Pvu*II restriction enzymes. The linearized plasmids were then treated with Klenow enzyme, deoxyadenosine 5'-[α-^{32}P]triphosphate and thymidine 5'-[α-^{32}P]triphosphate for 3' labeling. The reactions were loaded onto a 7% nondenaturing polyacrylamide gel. The desired bands were visualized by autoradiography and isolated. Chemical sequencing reactions were done according to published methods.

Quantitative DNase I footprinting.[59] DNase I footprinting reactions were carried out as previously described. Photostimulable storage phosphorimaging plates (Storage Phosphor Screen from Molecular Dynamics) were pressed flat against gel samples and exposed for 12–16 hours. Imaging of Storage Phosphor screens was accomplished on a Molecular Dynamics 425E PhosphorImager and the data analyzed using ImageQuant v. 3.2 software.

Binding energetics. Quantitative DNase I footprint titration experiments[59] (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl$_2$, 5 mM CaCl$_2$, pH 7.0, 22 °C) were performed on the 3'-^{32}P end labeled 270 bp *Eco*RI/*Pvu*II restriction fragment from pATK1 and the 3'-^{32}P end labeled 261 bp *Eco*RI/*Pvu*II restriction fragment from pATK2. Equilibrium association constants for polyamides **1a**, **2a**, and **3a** on the designed binding sites were determined by calculating a fractional saturation value at the site, for each polyamide concentration, and fitting the data to a modified Hill equation.

Cell uptake studies. Done with the help of Tim Best and Ben Edelson. Polyamide/cell incubations were performed by adding 150 μL cells into culture dishes equipped with glass bottoms for direct imaging. Incubations were done in medium containing 1 μM polyamide at 37 °C for 10–14 hours. Imaging was then performed with a Zeiss LSM 5 Pascal inverted laser scanning microscope using 488 nm laser excitation with a standard fluorescein filter set.

References

1. Darnell, J. E., *Nature Reviews Cancer*, **2002,** *2*, 740–749.

2. Pandolfi, P. P., *Oncogene*, **2001,** *20*, 3116–3127.

3. Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology*, **2003,** *13*, 284–299.

4. deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society*, **1997,** *119*, 7909–7916.

5. Mrksich, M.; Parks, M. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1994,** *116*, 7983–7988.

6. Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature*, **1996,** *382*, 559–561.

7. Bailly, C.; Chaires, J. B., *Bioconjugate Chemistry*, **1998,** *9*, 513–538.

8. Faria, M.; Giovannangeli, C., *Journal Of Gene Medicine*, **2001,** *3*, 299–310.

9. Reddy, B. S. P.; Sharma, S. K.; Lown, J. W., *Current Medicinal Chemistry*, **2001,** *8*, 475–508.

10. Kelly, J. J.; Baird, E. E.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **1996,** *93*, 6981–6985.

11. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nature Structural Biology*, **1998,** *5*, 104–109.

12. Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Chemistry-a European Journal*, **1997,** *3*, 1600–1607.

13. Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1998,** *120*, 3534–3535.

14. Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6160–6166.

15. Urbach, A. R.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2001,** *98*, 4343–4348.

16. Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., *Journal of Molecular Biology*, **2002,** *320*, 55–71.

17. Herman, D. M.; Baird, E. E.; Dervan, P. B., *Chemistry-a European Journal*, **1999,** *5*, 975–983.

18. Kers, I.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 3339–3349.

19. Weyermann, P.; Dervan, P. B., *Journal of the American Chemical Society*, **2002,** *124*, 6872–6878.

20. Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2003,** *100*, 12063–12068.

21. Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B., *Nucleic Acids Research*, **2004,** *32*, 2802–2818.

22. Goodsell, D. S.; Olson, A. J., *Annual Review of Biophysics and Biomolecular Structure*, **2000,** *29*, 105–153.

23. Mikhailov, V. S., *Molecular Biology*, **1999,** *33*, 498–510.

24. Bridson, P. K.; Orgel, L. E., *Journal of Molecular Biology*, **1980,** *144*, 567–577.

25. Bruick, R. K.; Dawson, P. E.; Kent, S. B.; Usman, N.; Joyce, G. F., *Chemistry & Biology*, **1996,** *3*, 49–56.

26. Calderone, C. T.; Puckett, J. W.; Gartner, Z. J.; Liu, D. R., *Angewandte Chemie-International Edition*, **2002,** *41*, 4104–4108.

27. Dolinnaya, N. G.; Sokolova, N. I.; Gryaznova, O. I.; Shabarova, Z. A., *Nucleic Acids Research*, **1988,** *16*, 3721–3738.

28. Ferris, J. P.; Huang, C. H.; Hagan, W. J., *Nucleosides & Nucleotides*, **1989,** *8*, 407–414.

29. Fujimoto, K.; Matsuda, S.; Takahashi, N.; Saito, I., *Journal of the American Chemical Society*, **2000,** *122*, 5646–5647.

30. Gartner, Z. J.; Grubina, R.; Calderone, C. T.; Liu, D. R., *Angewandte Chemie-International Edition*, **2003,** *42*, 1370–1375.

31. Gartner, Z. J.; Kanan, M. W.; Liu, D. R., *Journal of the American Chemical Society*, **2002,** *124*, 10304–10306.

32. Gartner, Z. J.; Kanan, M. W.; Liu, D. R., *Angewandte Chemie-International Edition*, **2002,** *41*, 1796–+.

33. Gartner, Z. J.; Liu, D. R., *Journal of the American Chemical Society*, **2001,** *123*, 6961–6963.

34. Kanaya, E.; Yanagawa, H., *Biochemistry*, **1986,** *25*, 7423–7430.

35. Lohrmann, R.; Bridson, P. K.; Bridson, P. K.; Orgel, L. E., *Science*, **1980,** *208*, 1464–1465.

36. Lohrmann, R.; Orgel, L. E., *Journal of Molecular Biology*, **1980,** *142*, 555–567.

37. Luo, P. Z.; Leitzel, J. C.; Zhan, Z. Y. J.; Lynn, D. G., *Journal of the American Chemical Society*, **1998,** *120*, 3019–3031.

38. Naylor, R.; Gilham, P. T., *Biochemistry*, **1966,** *5*, 2722–2789.

39. Niemeyer, C. M., *Current Opinion in Chemical Biology*, **2000,** *4*, 609–618.

40. Orgel, L. E.; Lohrmann, R., *Accounts of Chemical Research*, **1974,** *7*, 368–377.

41. Sokolova, N. I.; Ashirbekova, D. T.; Dolinnaya, N. G.; Shabarova, Z. A., *Febs Letters*, **1988,** *232*, 153–155.

42. Summerer, D.; Marx, A., *Angewandte Chemie-International Edition*, **2002,** *41*, 89–90.

43. Xu, Y. Z.; Karalkar, N. B.; Kool, E. T., *Nature Biotechnology*, **2001,** *19*, 148–152.

44. Xu, Y. Z.; Kool, E. T., *Journal of the American Chemical Society*, **2000,** *122*, 9040–9041.

45. Ye, J. D.; Gat, Y.; Lynn, D. G., *Angewandte Chemie-International Edition*, **2000,** *39*, 3641–3643.

46. Luebke, K. J.; Dervan, P. B., *Journal of the American Chemical Society*, **1989,** *111*, 8733–8735.

47. Huisgen, R., *Profiles, Pathways, and Dreams*. American Chemical Society: Washington, DC, 1994.

48. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., *Angewandte Chemie-International Edition*, **2001,** *40*, 2004–2006.

49. Lewis, W. G.; Green, L. G.; Grynszpan, F.; Radic, Z.; Carlier, P. R.; Taylor, P.; Finn, M. G.; Sharpless, K. B., *Angewandte Chemie-International Edition*, **2002,** *41*, 1053–1356.

50. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G., *Journal of the American Chemical Society*, **2003,** *125*, 3192–3193.

51. Warrener, R. N.; Butler, D. N.; Margetic, D.; Pfeffer, F. M.; Russell, R. A., *Tetrahedron Letters*, **2000,** *41*, 4671–4675.

52. Huisgen, R., *1,3-Dipolar Cycloaddition Chemistry*. Wiley-Interscience: New York, 1984; Chapter 1.

53. Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6141–6146.

54. Brunton, S. A.; Jones, K., *Journal of the Chemical Society-Perkin Transactions 1*, **2000**, 763–768.

55. Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 2767–2774.

56. Saxon, E.; Bertozzi, C. R., *Science*, **2000,** *287*, 2007–2010.

57. Fazio, F.; Bryan, M. C.; Blixt, O.; Paulson, J. C.; Wong, C. H., *Journal of the American Chemical Society*, **2002,** *124*, 14397–14402.

58. Hlasta, D. J.; Ackerman, J. H., *Journal of Organic Chemistry*, **1994,** *59*, 6184–6189.

59. Trauger, J. W.; Dervan, P. B., Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. In *Drug-Nucleic Acid Interactions*, 2001; Vol. 340, pp 450–466.

60. Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1999,** *121*, 1113–1120.

61. Baliga, R.; Baird, E. E.; Herman, D. M.; Melander, C.; Dervan, P. B.; Crothers, D. M., *Biochemistry*, **2001,** *40*, 3–8.

62. Gygi, M. P.; Ferguson, M. D.; Mefford, H. C.; Lund, K. P.; O'Day, C.; Zhou, P.; Friedman, C.; van den Engh, G.; Stolowitz, M. L.; Trask, B. J., *Nucleic Acids Research*, **2002,** *30*, 2790–2799.

63. Edayathumangalam, R. S.; Weyermann, P.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Proceedings of the National Academy of Sciences of the United States of America*, **2004,** *101*, 6864–6869.

64. Suto, R. K.; Edayathumangalam, R. S.; White, C. L.; Melander, C.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Journal of Molecular Biology*, **2003,** *326*, 371–380.

Chapter 3

Turn-to-Turn Dimerizations of Hairpin Polyamides on Duplex DNA Templates and on NCP Templates

Abstract

Double-helical DNA accelerates the rate of ligation of two six-ring hairpin polyamides, which bind adjacent sites in the minor groove via Michael additions and 1,3-dipolar cycloadditions to form turn-to-turn dimers. The rate of the templated reaction is dependent on DNA sequence as well as on the distance between the hairpin binding sites. Turn-to-turn ligation is also being explored across the "supergroove" of a nucleosome core particle (NCP). Progress towards a fluorescent readout of templated dimerization is also reported.

Introduction.

Hairpin polyamides have proven useful in binding predetermined sequences of DNA in a sequence-specific fashion.[1-3] Furthermore, this class of molecules has been shown to inhibit the binding of many DNA-binding proteins *in vitro*.[4-11] For the *in vivo* use of these molecules to be realized, they must be able to transverse the cellular and nuclear membranes of live cells in order to reach their target DNA.[12, 13]

For applications in gene regulation within biological systems, binding-site size may be critical because longer sequences should occur less frequently in a gigabase-sized genome. For this reason, the design of ligands capable of targeting >10 base pairs of DNA remains an important goal in the area of polyamide design.[3, 14-16] One approach to increase polyamide binding-site size has been to covalently link two hairpin modules to form hairpin dimers. Dimers linked both "turn-to-tail" and "turn-to-turn" have excellent affinity and specificity to DNA sequences up to 10 bp in length.[17-19] Though likely satisfying the DNA-binding criteria to target unique sequences within genomic DNA, hairpin dimers do not possess the favorable cell and nuclear uptake properties of smaller hairpins, presumably due to size and shape.[12, 13]

In the preceding chapter of this thesis, we demonstrated that duplex DNA can template the formation of "turn-to-tail" tandem hairpin dimers from azide- and alkyne-functionalized hairpin precursors via a 1,3-dipolar cycloaddition reaction.[20] These data showed that tandem-type polyamides capable of targeting sequences of DNA >10 bp could be created from small hairpin starting materials that possess more favorable nuclear localization properties than their dimer products, which will be necessary for *in vivo*

applications. It was found, however, that while alkyne-functionalized polyamides were able to localize to live cell nuclei, polyamides functionalized with azide moieties on the C-terminus were excluded from the nuclei in most cell lines tested. While the azide moiety itself appears to be a negative determinant of nuclear localization, it was found that when the azide is switched from the C-terminus to the internal γ-aminobutyric acid "turn" position, a significant increase in nuclear localization was observed.
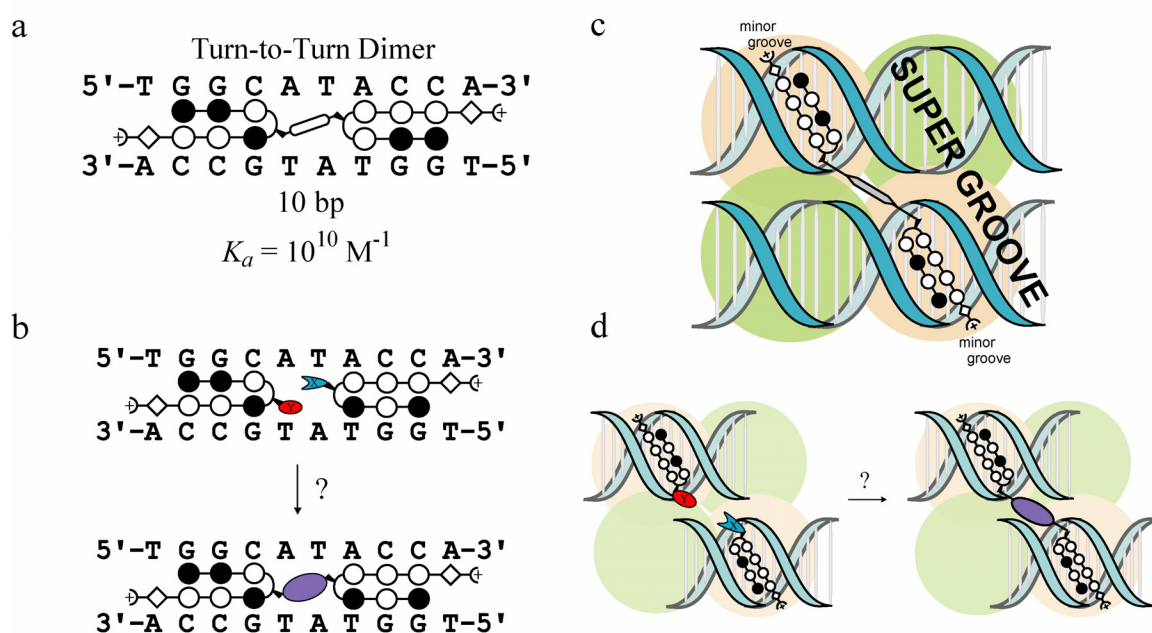
Figure 3.1. **a.** Schematic representation of a turn-to-turn hairpin dimer found to bind the 10 base pair sequence shown with 10^{10} binding affinity. **b.** Schematic representation of DNA-templated turn-to-turn dimerization; polyamides functionalized with complementary reactive groups "x" (blue shape) and "y" (red shape) bind DNA. The close proximity of x and y causes a covalent linkage to form between the two polyamides (purple shape). **c.** Schematic representation of "supergroove" recognition by turn-to-turn polyamide dimers. **d.** Schematic representation of how DNA bound to the NCP might template the dimerization of hairpin polyamides bound to a "supergroove." Here, polyamide binding sites located 80 linear base pairs apart are placed in close proximity upon winding around the nucleosome core proteins. In each figure, pyrrole is represented by open circles, imidazole by closed circles.

Turn-to-turn hairpin dimers have been shown to bind the minor groove of B-form DNA[19] as well as the "supergroove" created when DNA is packaged into nucleosome core particles (Figure 3.1).[21] Because of these successes, coupled with our cell uptake

data, we ask whether DNA is able to template the dimerization of hairpin polyamides in a "turn-to-turn" fashion. In the turn-to-turn orientation, hairpin polyamide starting materials will be functionalized only at the turn position, which, in our previous work was shown to be optimal for nuclear localization. Perhaps this architecture will produce molecules capable of entering the nuclei of live cells, and using the genomic DNA to template the formation of "turn-to-turn" hairpin dimers that target large, genomically unique sequences of DNA (Figure 3.1).

Experimental Design.

Because our previous work indicated that the azide moiety is a negative determinant of nuclear uptake,[22] the scientific literature was searched in order to find other water-compatible organic reactions that do not require additional cofactors or catalysts.[23, 24] The 1,3-dipolar cycloaddition between an azide and an alkyne is not the only reaction that fits the above criteria. One can envision templated dimerizations through a wide variety of reactions such as the Diels-Alder reaction, Michael addition, S_N2 nucleophilic substitution, or Wittig reaction. Any of these reactions may produce templated dimers from functionalized hairpin polyamides capable of entering live cell nuclei (Figure 3.2).

Because turn-to-turn dimers have been shown to bind two distinct DNA architectures, we are interested in whether the two architectures will have different templating properties. The minor groove of B-form DNA is narrow, and will not accommodate steric bulk.[25-27] Furthermore, each single base pair rise offers approximately 5 Å across which the "turn-to-turn" templated reaction may occur.

Conversely, the gap between the two gyres of DNA on the nucleosome core particle is deep and wide, and may be able to accommodate bulky groups.[28] Structural data shows that the gap between two 8-ring polyamides bound to a nucleosomal "supergroove" is 11 Å. Because of their different molecular structures, these two DNA architectures should have different templating properties. We begin our studies examining the templating properties of linear, duplex DNA.
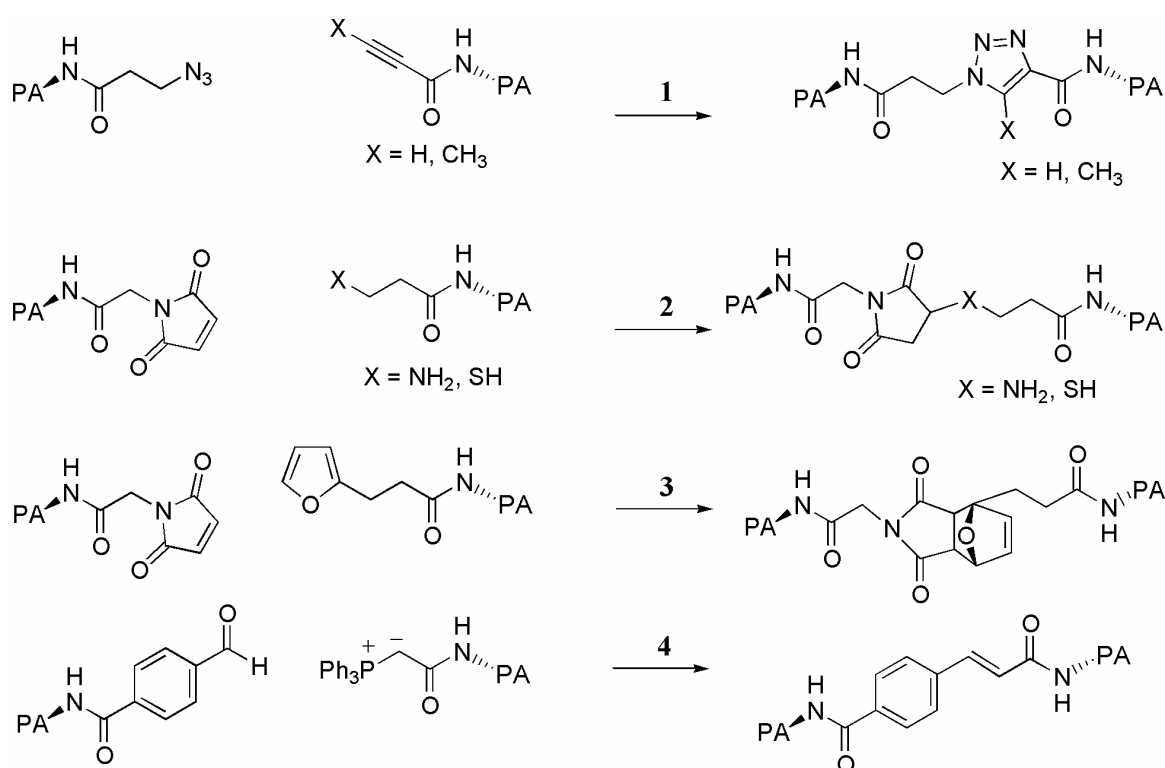
Figure 3.2. Structures of several dimerization reactive partners and the putative linkers that form upon dimerization. **1**. 1,3-dipolar cycloaddition forms a triazole linker. **2**. Michael addition forms a cycle-containing linker. **3**. Diels-Alder reaction forms a linker containing a bulky fused ring system. **4**. Wittig reaction forms an unsaturated C-C bond-containing linker.

Polyamide Design and Synthesis.

Figure 3.3 shows the chemical structures of the functionalized hairpin polyamides synthesized for this study. Unfunctionalized polyamides **1** and **2** were

synthesized on solid support[29, 30] and liberated from resin using dimethylamino propylamine (Dp), leaving a single free amine on the turn residue. Polyamides were then functionalized either with N-hydroxysuccinimidyl esters of the functional groups, or by *in situ* activation of carboxylic acids with PyBOP.
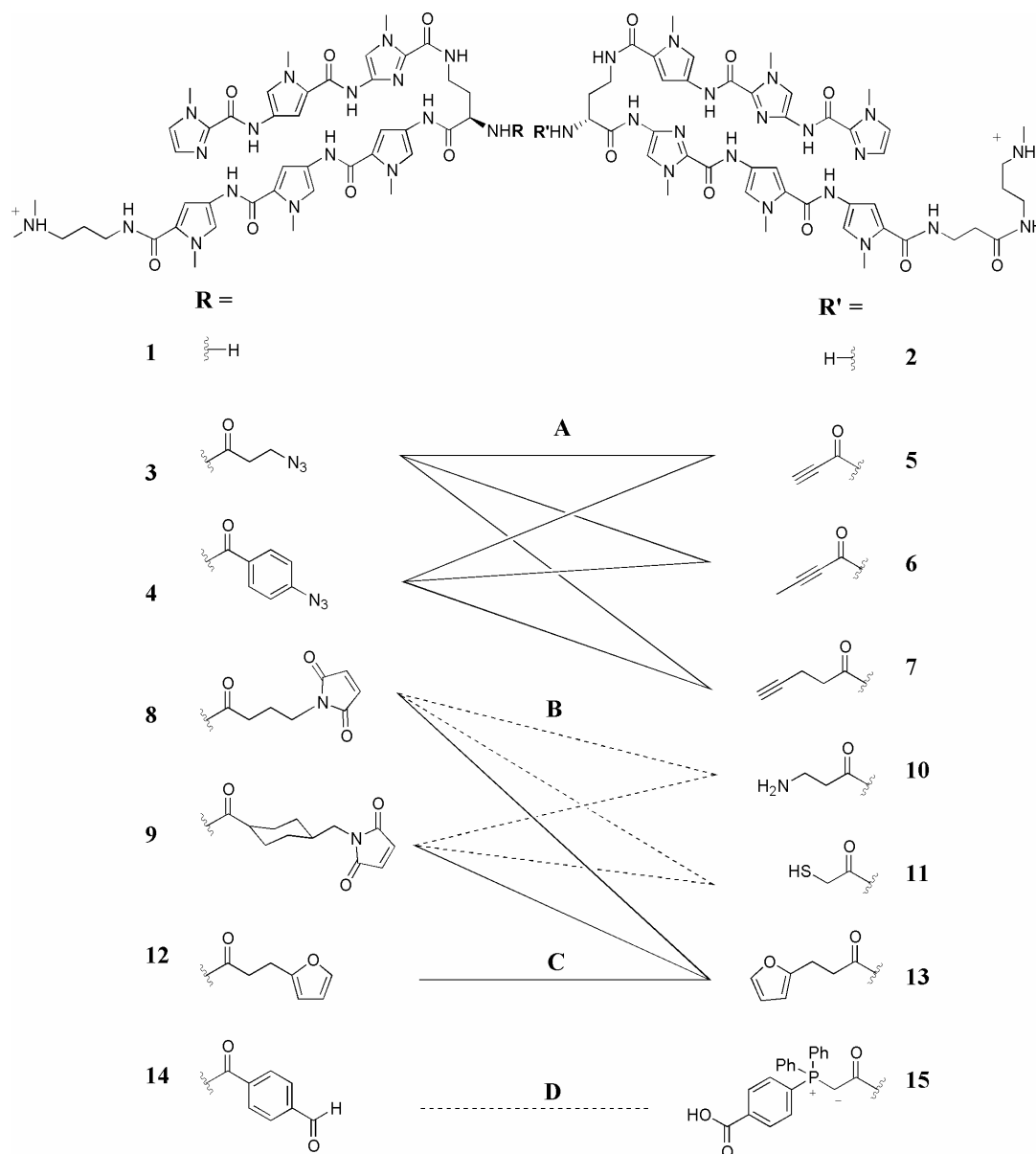
Figure 3.3. Chemical structures of the polyamides synthesized for turn-to-turn dimerizations on duplex DNA. Compounds **1** and **2** were synthesized on solid support. Molecules **3–15** were synthesized by amide bond–formation between the turn amines on **1** and **2** and activated esters containing each of the side chains. Lines between columns indicate potential reactive partners via dipolar cycloadditions (**A**), Michael additions (**B**), Diels-Alder reactions (**C**), or the Wittig reaction (**D**).

Several polyamides were created with identical reactive groups, but with side chains of varying length and steric bulk. These differences were designed into the ligands in order to exploit the molecular differences between linear and nucleosomal DNA templates, thereby leading to molecules that may be capable of specifically forming dimer product on only one of the two architectures (e.g., polyamides **5** and **6** each contain the maleimide reactive group. However, **5** contains a small alkyl linker while **6** incorporates a bulky cyclohexyl linker. Perhaps **5** will react well within the confines of the narrow minor groove while **6** will be sterically excluded from linear DNA, reacting only across the spacious "supergroove").

DNA-Templated Dimerization on Linear, Duplex DNA.

Initial studies were designed to test the ability of match and mismatch linear DNA to template the various turn-to-turn dimerizations. It is anticipated that the spacing

GGGGTATGGCAACACTATGGGG GGGGTATGGCATATACACTATGGGG
CCCCATACCGTTGTGATACCCC CCCCATACCGTATATGTGATACCCC
   (A) 10 bp match   (D) 13 bp match

GGGGTATGGCATACACTATGGGG
CCCCATACCGTATGTGATACCCC
   (B) 11 bp match

GGGGTATGGCATTACACTATGGGG GGGGTATGGGAAGACTATGGGG
CCCCATACCGTAATGTGATACCCC CCCCATACCCTTCTGATACCCC
   (C) 12 bp match   (E) 10 bp mismatch
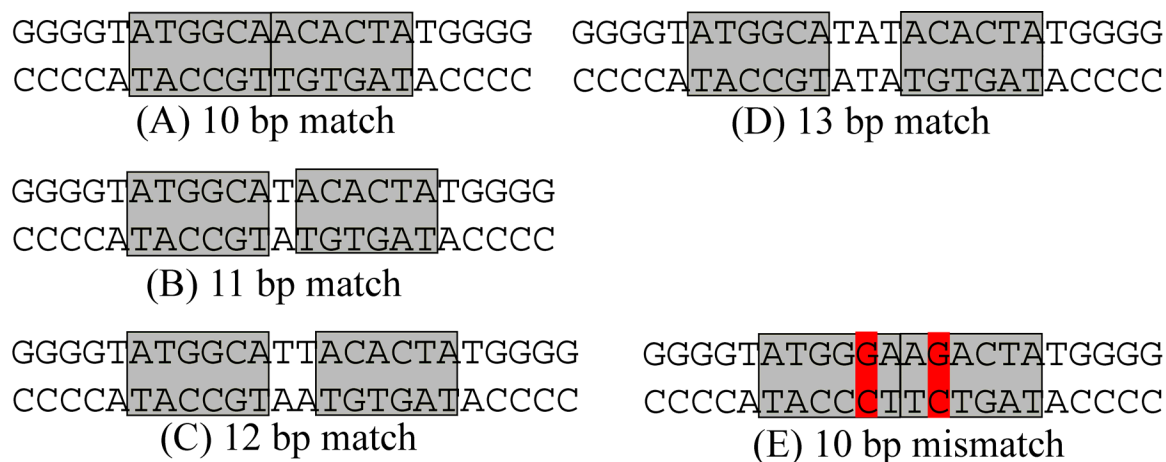
Figure 3.4. Sequences of the duplex oligonucleotides containing the match sites for each of the two hairpin polyamides separated by zero, one, two, or three base pairs (**A–D**, respectively). Duplex **E** contains a formal polyamide mismatch under each of the two hairpin binding sites (which are separated by zero base pairs). Sequences are listed with the top strand oriented in the 5'→3' direction. Hairpin binding sites are boxed in gray. Mismatch bases are highlighted in red.

between the hairpin binding sites will influence the rates of the templated reactions. To assess the optimal polyamide separation distance, the duplex templates 5'-GGGGT**AGGCATCACA**TGGGG-3' (**A**), 5'-GGGGT**AGGCAT**T**CACA**TGGGG-3'(**B**), 5'-GGGGT**AGGCA**TT**CACA**TGGGG-3 (**C**), 5'-GGGGT**AGGCA**TAT**TCA-CA**TGGGG-3' (**D**), were synthesized (Figure 3.4). Each template contains five base pair match sites for each of the hairpin polyamides **1** and **2**, separated by zero, one, two, or three base pairs, respectively.

Initially, each reaction was screened for high templated:nontemplated yield ratios. Reactions were performed with equal concentrations of each hairpin polyamide and DNA (2 mM Tris-HCl, 2 mM KCl, 2 mM MgCl$_2$, 1 mM CaCl$_2$, pH 7.0, 37 ºC. If a nucleophilic amine was involved, reactions were performed in 2 mM Tris-HCl, 2 mM KCl, 2 mM MgCl$_2$, 1 mM CaCl$_2$, pH 8.2, 37 ºC. If a nucleophilic sulfur was involved, reactions were performed in 2 mM Tris-HCl, 2 mM KCl, 2 mM MgCl$_2$, 1 mM CaCl$_2$, pH 7.4, 37 ºC). Reactions were monitored by analytical reversed-phase HPLC. MALDI-TOF MS was used to verify product formation.

Data for the initial templated reaction screen are summarized in Table 3.1. Neither the Wittig (polyamides **14 + 15**) nor the Diels-Alder (polyamides **8 + 13**, **9 + 13**, and **12 + 13**) reactions form detectable amounts of product on any of the four match templates at 1 μM concentrations in a 24-hour period. The Michael additions with the sterically small maleimide **8** (polyamides **8 + 10**, and **8 + 11**) exhibit product formation on both the 10- and 11-base pair templates (**A** and **B**) at 1 μM concentrations, with the reactions proceeding to 7% and 13% (**8 + 10**), and 23% and 26% (**8 + 11**), respectively, in 24 hours. However, under non-templated conditions, products also form in

approximately 5% (**8 + 10**) and 15% (**8 + 11**) yields in 24 hours. Michael additions with

the bulky maleimide **9** (polyamides **9 + 10**, and **9 + 11**) were also observed in the absence

of template, achieving 5% (**9 + 10**) and 15% (**9 + 11**) yields in 24 hours. However, all

four DNA templates produced no reaction between these polyamides in 24 hours.

Table 3.1. Tandem product formation after 24 hours.[a, b]

	5'-AGGCATCACT-3' A	5'-AGGCATTCACT-3' B	5'-AGGCATTTCACT-3' C	5'-AGGCATATTCACT-3' D	No DNA
3 + 5	78%	37%	---	---	---
3 + 6	35%	---	---	---	---
3 + 7	5%	4%	---	---	---
4 + 5	---	---	---	---	---
4 + 6	---	---	---	---	---
4 + 7	---	---	---	---	---
8 + 10	7%	13%	---	---	5%
8 + 11	23%	26%	---	---	15%
9 + 10	---	---	---	---	5%
9 + 11	---	---	---	---	15%
8 + 13	---	---	---	---	---
9 + 13	---	---	---	---	---
12 + 13	---	---	---	---	---
14 +15	---	---	---	---	---

[a] The assays were carried out at 37 °C at pH 7.0 in the presence of 3 mM Tris-HCl, 3 mM KCl, 3 mM MgCl$_2$, and 1.7 mM CaCl$_2$, 1 μM each polyamide, 1μM DNA. [b] Yield was quantitated from analytical HPLC traces taken at 24 hours after reaction initiation. Dashed lines indicate no product detected.

1,3 dipolar cycloadditions with the sterically bulky azide **4** (polyamides **4 + 5**,

4 + 6, and **4 + 7**) produce no dimer product in the absence of template, nor with any of

the four duplex templates.

The 1,3 dipolar cycloadditions with the sterically small azide **3** (polyamides **3** + **5**, **3** + **6**, and **3** + **7**) produce no dimer product under non-templated conditions at 1 μM concentrations. When **3** and **7** are incubated with template **A** or **B** at 1 μM concentrations, product is detected after 24 hours (37 ºC). Quantitation of HPLC data indicates that the reaction proceeds in approximately 5% yield in 24 hours on each template. No reaction is observed on templates **C** or **D**.

When polyamides **3** and **5** are incubated with templates **A** or **B** at 1 μM concentrations, product is detected after 45 minutes (37 ºC). Quantitation of HPLC traces indicate that, after 24 hours, template **A** yields 78% product while template **B** yields 37% dimer product. No reaction is observed on templates **C** and **D**.

When polyamides **3** and **6** are incubated with template **A** at 1 μM concentrations, dimer product is observed in 35% yield after 24 hours. No reaction is observed on templates **B**, **C**, or **D**.

Reaction Order (Table 3.2).

Table 3.2. Pseudozero-order rate constants for the 1,3-dipolar cycloaddition between polyamides **3** and **5**.[a,b]

3 + 5	5'-AGG**CATCACT**-3' **A**	5'-AGG**CATTCACT**-3' **B**	5'-AGG**GATGACT**-3' **E**	No DNA
1 μM	25,400	9360	8270	≤ 1
500 nM	25,300	11700	2520	≤ 1

[a] The assays were carried out at 37 °C at pH 7.0 in the presence of 3 mM Tris-HCl, 3 mM KCl, 3 mM MgCl$_2$, and 1.7 mM CaCl$_2$, 1 μM each polyamide, 1μM DNA. [b] Rate constants were calculated from the slope of the first four data points (4.5 hours) when % completion is plotted as a function of time.

Because the 1,3-dipolar cycloaddition between **3** and **5** possesses the most favorable templated yields (while producing no dimer product under nontemplated

conditions), the kinetics of this reaction were studied in depth. Using analytical HPLC, reactions were quenched and quantitated every 1.5 hours for 12 hours. Reaction on template **A** produces 16% dimer product in 1.5 hours, while reaction on template **B** produces 9% product in the same time (Figure 3.5).
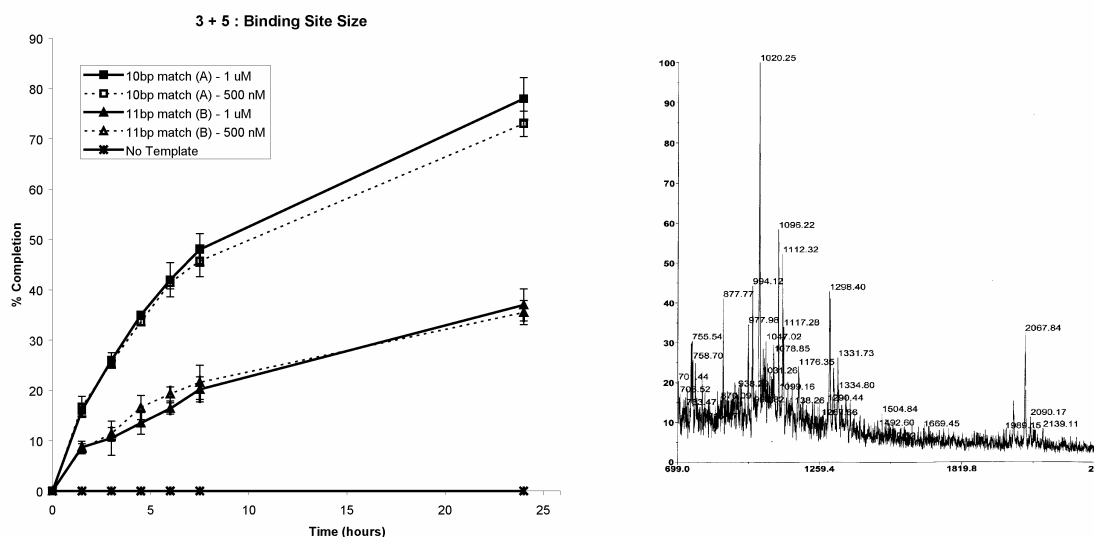
Figure 3.5. Left panel: Rate data for the 1,3-dipolar cycloaddition between azide-functionalized polyamide **3** and alkynyl amidate–functionalized polyamide **5** at 1 μM and 500 nM concentrations on templates **A** and **B**. Rates of product formation at the two different concentrations are identical, indicating pseudo zero-order kinetics. Right panel: MALDI-TOF mass spectrometry data taken from the reaction of **3** and **5** on duplex template **A** after 8 hours. Product mass is expected at 2067.3 $[M + H]^{+}$.

Similarly, when **3** and **5** are incubated with turn-to-turn templates **A** and **B** at 500 nM concentrations, the rates and yields of reactions on these match templates are identical to those at 1 μM. Thus, the reaction is independent of concentration, and thus, is a unimolecular process. By plotting product formation as a function of time, pseudozero-order rate constants can be obtained for these reactions. Template **A** increases the rate of dimer formation relative to the non-templated reaction approximately 25,000-fold. Likewise, template **B** increases the rate of dimer formation approximately 10,000-fold.

Template Mismatch Tolerance: 10 Base Pair Binding Site.

To assess if the cycloaddition reaction between **3** and **5** is sequence-specific with respect to the template, the duplex corresponding to 5'-GGGGT**AGG***G***AT***G***A**CATG-GGG-3 (**E**) was synthesized. This duplex contains the 10 base pair binding site (zero intervening bases) with a single mismatch (italicized) under each of the five base pair hairpin polyamide binding sites. The 10 base pair site was chosen for these studies because site-separation preference results indicate this site to be optimal for product formation.

When **3** and **5** are incubated with mismatch oligo **E** at 1 μM concentrations, product is formed approximately 4-fold slower than on template **A**, eventually achieving 25% yield after 24 hours. When **3** and **5** are incubated with **E** at 500 nM concentrations, the rate is reduced by an additional 4-fold versus the reaction on match template **A**, thereby achieving more than 10-fold match versus mismatch selectivity (Figure 3.6).
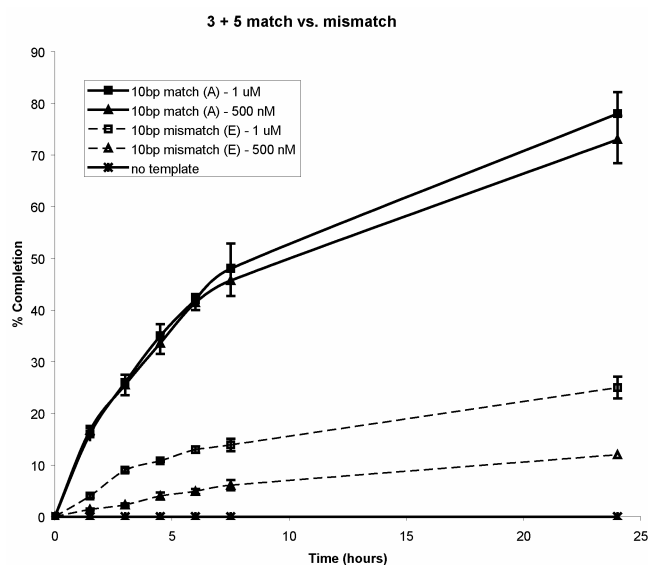

Figure 3.6. Rate data for the 1,3-dipolar cycloaddition between polyamides **3** and **5** on the 10 base pair match template (**A**) and the 10 base pair mismatch template (**E**) at 1 μM and 500 nM concentrations. At 500 nM, the reaction proceeds 10 times faster on the match template than on the mismatch template.

Towards "Supergroove"-Templated Turn-to-Turn Dimerizations.

Research has shown that the 146 and 147 base pair duplex DNAs derived from human α-satellite DNA forms well-positioned nucleosome core particles (NCP) with the histone proteins.[21, 28, 31] Furthermore, cocrystal structures were obtained of these NCPs and various polyamides. Strikingly, because of the palindromic nature of the DNA, a single polyamide was shown to bind two sites located 80 linear base pairs apart, yet juxtaposed in a single "supergroove" by the two gyres of DNA wound around the histone protein core.[28] Subsequent studies linked these two polyamides in a turn-to-turn fashion, forming a homodimeric nucleosome clamp.[21] This clamp was able to effectively prevent NCP melting by locking a full circle of DNA onto the NCP. These types of NCP clamps could have interesting uses for gene silencing in living systems. However, the clamps are branched oligomers of large size and, as such, are unable to translocate across cellular and nuclear membranes.[12] Perhaps the NCP can be used to template the formation of clamplike turn-to-turn dimers from cell-permeable hairpin polyamide pieces.

The templated reactions analyzed thus far on duplex DNA have all relied on two different reactive species. Thus, in order to pursue NCP-templated dimerizations, a single "supergroove" that has two different binding sites is needed (recall that the NCP clamp is a homodimer[21]). A survey of the sequences of each of the six supergrooves around the NCP crystal structure with α-satellite DNA does not yield sites optimal for polyamide binding. Additionally, information about polyamide binding to each of the other supergrooves does not contain structural data of the precise positioning of polyamides. Because of these factors, it was decided to mutate two base pairs of a single

site of the supergroove to which the NCP clamp binds, thereby mutating that supergroove

into a heterodimeric binding site (Figure 3.7). The site was chosen such that 6- or 8-ring
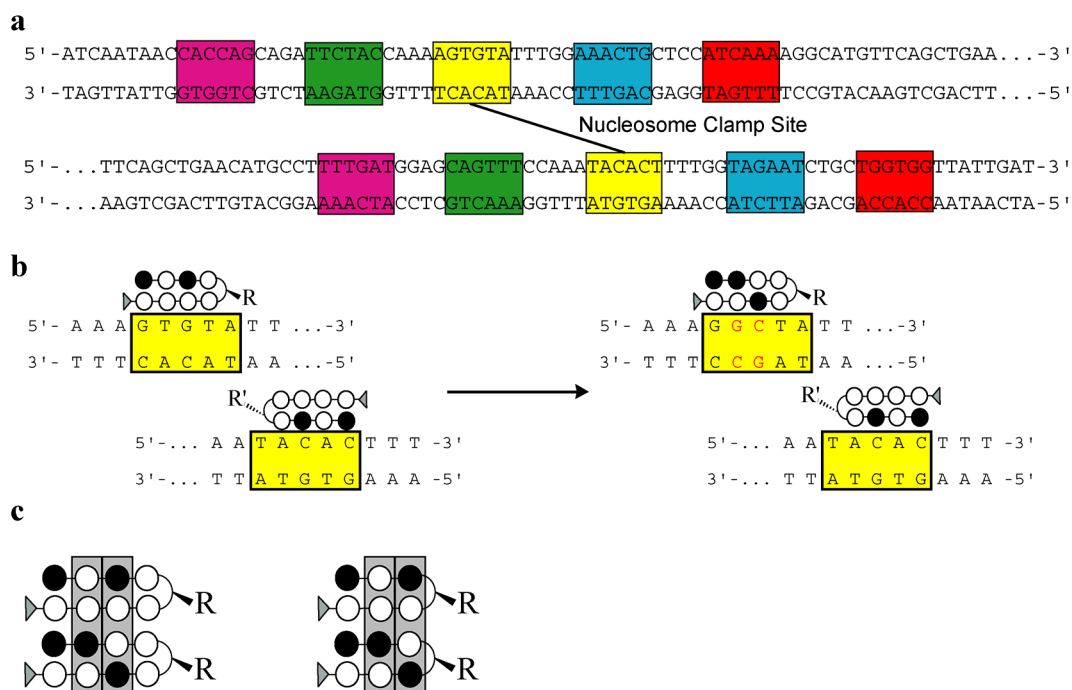
Figure 3.7. **a.** Sequence of the 146 base pair fragment of α-satellite DNA used for crystallographic studies with the nucleosome clamp. In those structures, the clamp was found to bind in the homodimeric "supergroove" highlighted in yellow. Each of the other four supergrooves on the NCP are highlighted in purple, green blue and red. **b.** Highlight of the sequences to which the nucleosome clamp was bound (left). At right is shown the two base pair mutation (in red) introduced so that the supergroove becomes heterodimeric. Located above each highlighted site is the polyamide designed to target that site. **c.** Illustration of the fact that both 8-ring and 6-ring hairpin polyamides can be designed for the new supergroove sequence (**b**, right). In each case, the two polyamides differ from each other by a double base pair mismatch (highlighted in gray).

hairpin polyamides could be designed to bind the sites. In each case, the two molecules

differ from each other by a double base pair mismatch. Additionally, the sites were

designed such that the 6-ring polyamides that match the sites are the same polyamides

used the duplex template study described above. Thus, new 6-ring polyamides will not

need to be synthesized for this study. One additional concern is that because nucleosome

positioning is very important to this study, it is hoped that this small perturbation does

not affect the character of this DNA that causes it to be well-positioned.

We will introduce the mutated bases by a combination of chemical synthesis and biochemical techniques. The schematic for creation of the 146 base pair fragment is outlined in Scheme 3.1. Briefly, three ~50mers of duplex DNA will be chemically synthesized. Each will have an overhanging 4 base pair region that is complementary to the strand to which it is to be ligated. T4 DNA ligase will then be used to stitch the three pieces together. Once small quantities of the full 146mer have been synthesized, polymerase chain reaction (PCR) will be used to obtain sufficient quantities for kinetic assays. With the DNA in hand, NCPs will be reconstituted and incubated with functionalized hairpin polyamides, and dimer product formation monitored by quantitation of analytical HPLC traces.
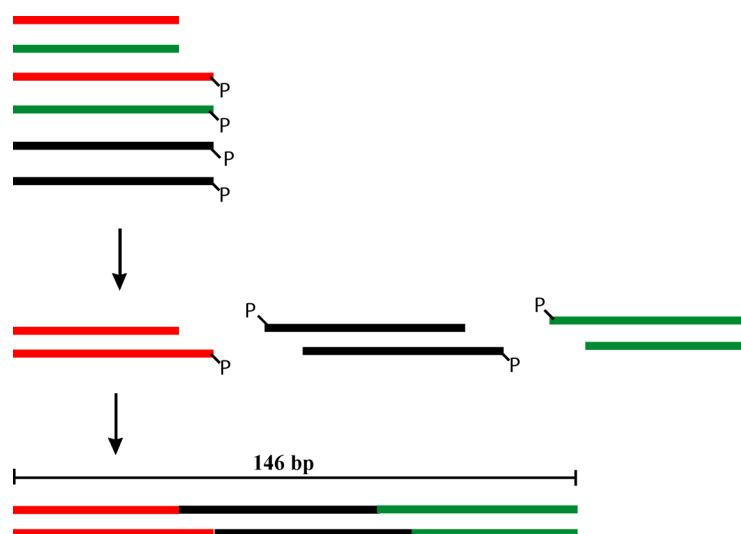
Scheme 3.1. Schematic of the semisynthesis of the 146 base pair fragment of DNA for NCP-mediated ligation studies. Six ~50mers will be chemically synthesized, four being 5'-phosphorylated (sequences represented by red, green, and black lines; complementary sequences are the same color.). Complementary sequences will be annealed. Sequences are designed such that the resulting duplexes have four base pair overhangs that are complementary to the overhangs on the duplexes to which they will be ligated (i.e., red overhang is complementary to the left black overhang; right black overhang is complementary to the green overhang). T4 DNA ligase will then be used to stitch the entire 146 base pair fragment together. PCR will be used to amplify the fragment to obtain amounts necessary for kinetic studies.

Profluorescent Azidocoumarins: Fluorescent Readout of Dimerizations.

Recently, researchers showed that installation of an azide into the 3- position of coumarins quenches fluorescence. Upon reaction of the azide with alkynes via a 1,3-dipolar cycloaddition, fluorescence is restored (Figure 3.8).[32] By functionalizing
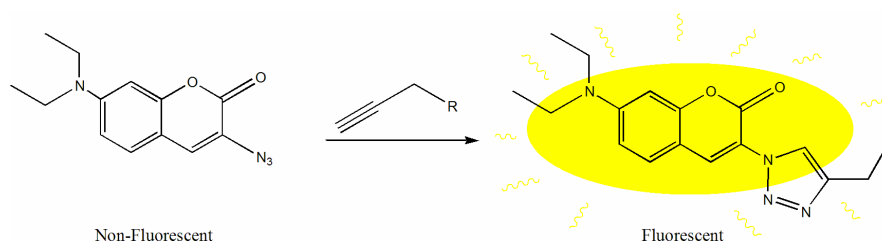
Figure 3.8. Representation of fluorescence rescue by 1,3-dipolar cycloadditions. 3-azidocoumarins are non-fluorescent because the lone pairs on the azide moiety are able to donate back into the ring and quench the excited state. Upon reaction with an alkyne, the electron-withdrawing triazole no longer quenches the excited state, and coumarin fluorescence is restored.

polyamides designed to the NCP "supergroove" with a 3-azidocoumarin and an alkyne, fluorescence rescue can be used as a readout for templated dimerization (Figure 3.9). Templated dimerizations can thus be monitored (even from the insides of living cells) in a non-invasive fashion via fluorescence microscopy.
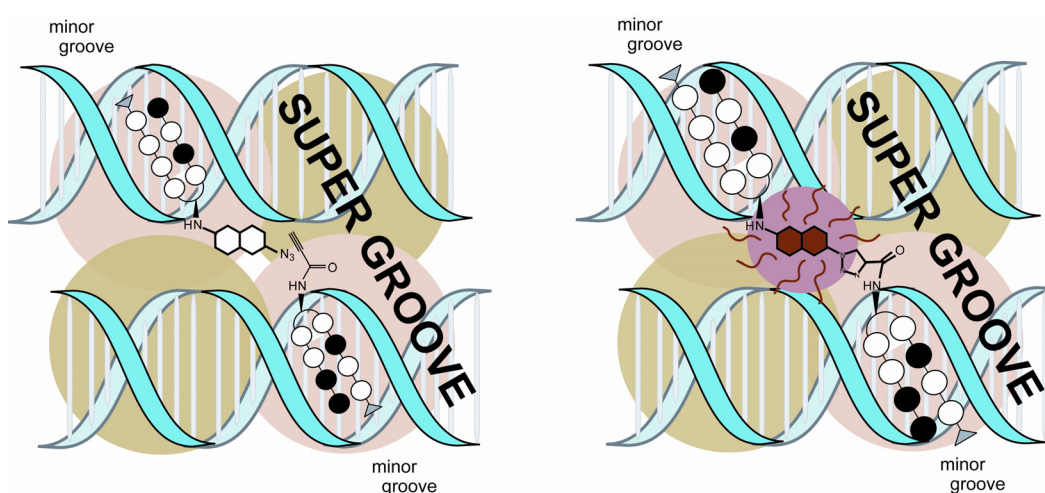
Figure 3.9. Schematic of fluorescence rescue by NCP-mediated polyamide dimerization. Azidocoumarin- and alkyne-functionalized polyamides bind to a "supergroove" (right), placing the reactive groups in close proximity. The templated cycloaddition then forms the fluorescent polyamide dimer product (left).
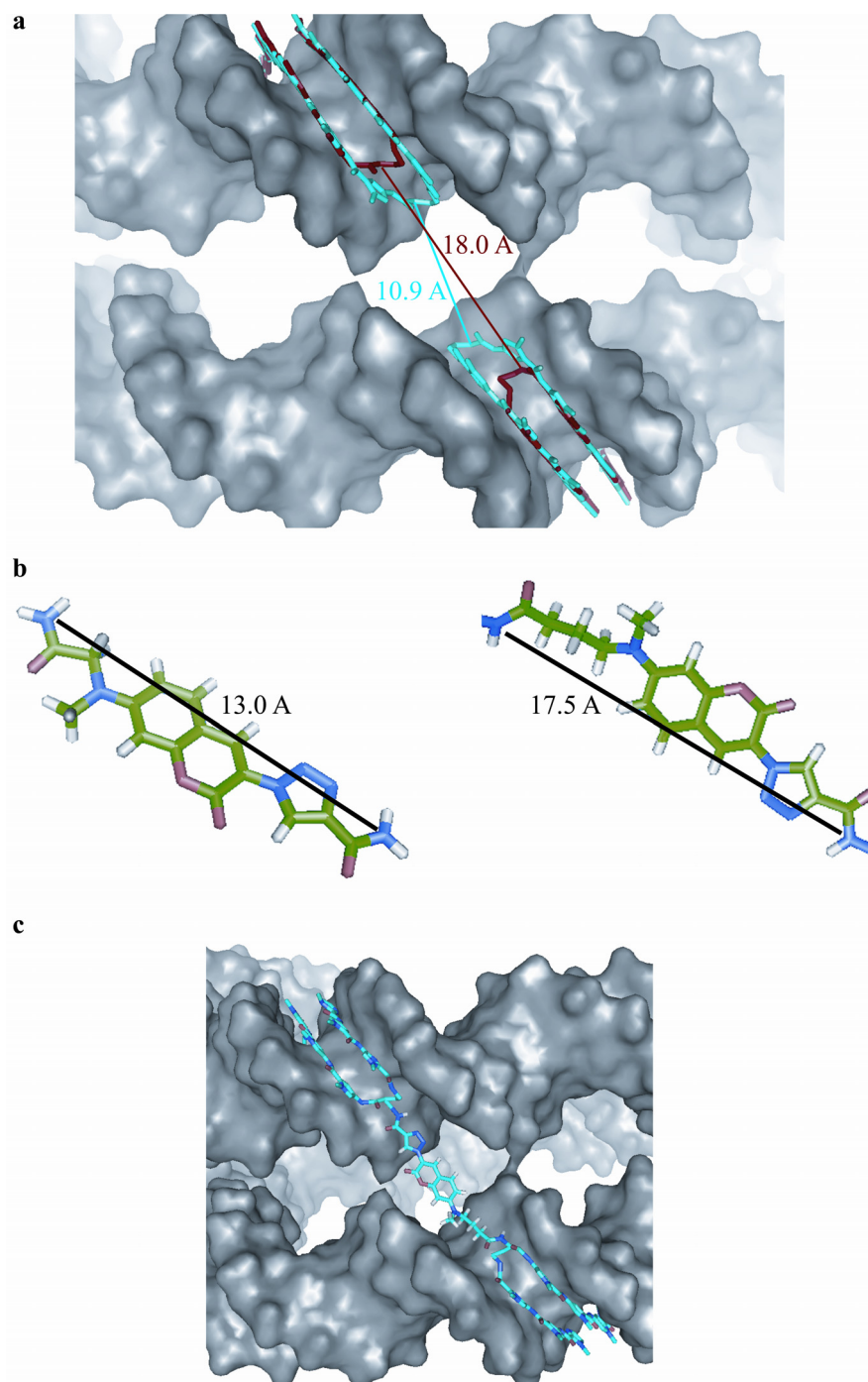
Figure 3.10. **a.** X ray crystal structure of 8-ring (blue) polyamides bound to the "supergroove" highlighting the 11 Å distance between hairpins. When 6-ring polyamides are modeled in (red), the distance between polyamides is increased to 18 Å. **b.** Energy-minimized models of triazole-coumarin linkers with 1 (left) and 3 (right) methylene spacers. Labeled atomic distances show these linkers to be good fits for the gap between hairpins bound to the NCP supergroove. Green = carbon, white = hydrogen, blue = nitrogen, red = oxygen. **c.** Model of the longer triazole coumarin linker forming a dimer between two 6-ring hairpin polyamides bound to a single supergroove.

In order to establish the feasibility of this scheme, molecular modeling was done to verify that the coumarin-triazole linker would appropriately span the gap across the two gyres of DNA on the NCP. A model of the triazole-coumarin linker was built in the Spartan ES software package and energy-minimized using an AM1 model, followed by *ab initio* calculations by means of the Hartree-Fock model and a 6-31G* basis set. As shown in Figure 3.10, the fully extended structure of the shortest triazole-coumarin spans 13 Å. This distance approximates the 10.99 Å distance between the two polyamide turn amines in the crystal structure. When 6-ring hairpin polyamides are substituted for the 8-ring polyamides on the crystal structure, the distance between turn amines increases to 17.8 Å. A triazole-coumarin linker formed from an alkynyl amide and an azidocoumarin functionalized with a three-carbon linker creates an approximately 18 Å linker. Thus, modeling shows that triazole-coumarins can be accommodated by the gap between hairpin biding sites within a single NCP supergroove.

Towards synthesis of functionalized 3-azidocoumarins.

Based on our previous templating work, it is clear that the kinetics of templated 1,3-dipolar cycloadditions are most favorable when alkynes directly conjugated to an electron-withdrawing group are used. While previous literature shows alkyl alkynes capable of fluorescence rescue, it does not offer any information on whether alkynyl amidates can be used with 3-azidocoumarins to rescue fluorescence. As initial controls, in separate reactions, 7-diethylamino 3-azidocoumarin was combined with **5** and **7**, and allowed to react at room temperature under copper-mediated conditions.[33] After two hours, reactions were deemed complete by HPLC. Solutions of the starting material and

products were irradiated with 365 nm light. As shown qualitatively in Figure 3.11, the products are intensely fluorescent with respect to the 3-azidocoumarin starting material. Thus, alkynyl amidates can be used to rescue the fluorescence of 3-azidocoumarins via a 1,3-dipoar cycloaddition.
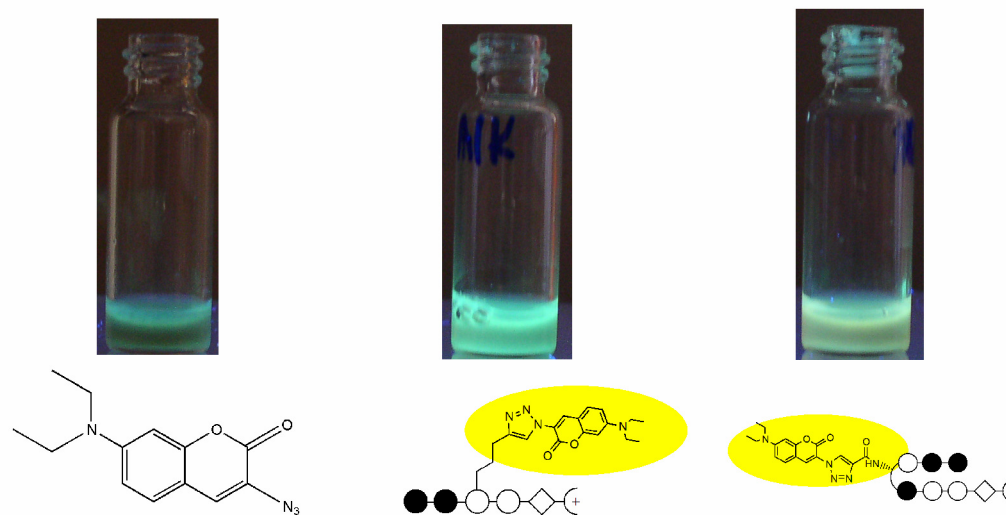
Figure 3.11. Aqueous solutions of azidocoumarin (left), alkyl triazole coumarin (center), and alkynyl amidate triazole coumarin (right) illuminated with 365 nm light. As shown, the azidocoumarin is relatively non-fluorescent, while the two triazole coumarins are intensely fluorescent.

With the necessary controls performed, our attention was turned towards synthesis of a 3-azidocoumarin that could be conjugated to a polyamide. Figure 3.12 shows our planned synthetic scheme and our progress towards this molecule. The decision was made to create azidocoumarins with a tertiary amine in the 7 position, since these molecules have been shown to be the most intensely fluorescent.[32] To begin the synthesis, m-aminophenol is mono-N-methylated by stepwise formylation and reduction to yield compound **17**.[34] The secondary amine is then alkylated with methylbromoacetate to yield tertiary amine **18** with a protected handle for linkage to a polyamide.[35] Formylation of the aromatic ring is accomplished under Vilsmier conditions to yield
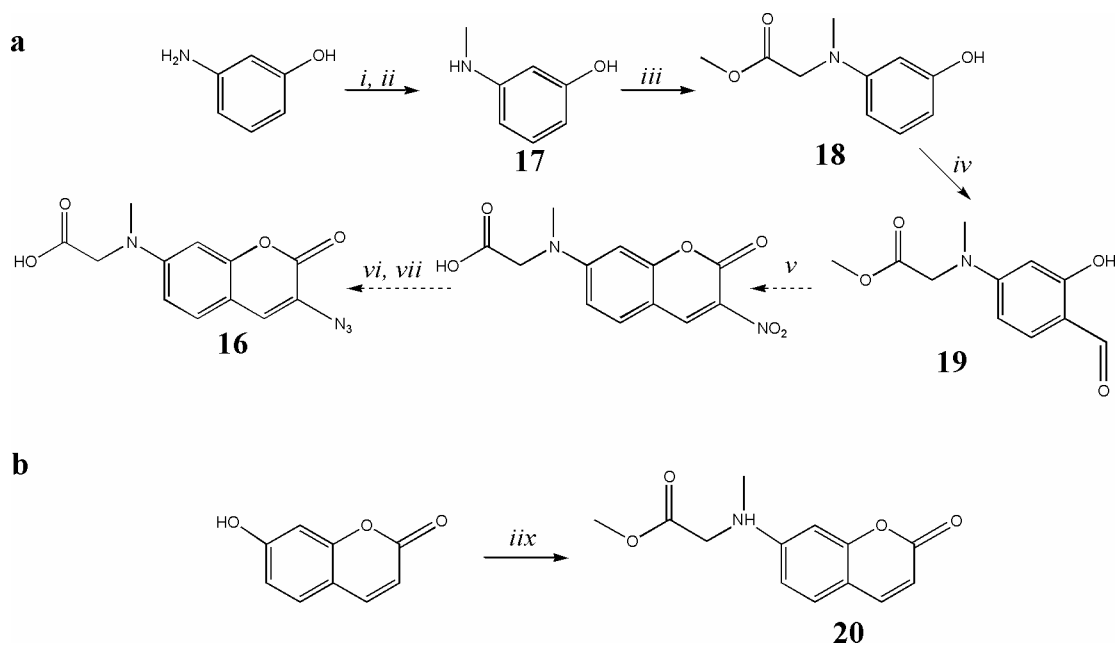
Figure 3.12. **a**. Synthetic scheme for an azidocoumarin that can be attached to a polyamide. i) ethyl formate, reflux, ii) BH$_3$-THF, THF (60% for two steps), iii) methyl bromoacetate, 2,6 lutidene, DMF (85%), iv) POCl$_3$, DMF (35%), v) nitro ethyl acetate, pyridine, acetic acid, toluene, vi) SnCl$_2$, HCl, vii) NaNO$_2$, KOAc, NaN$_3$. **b**. Buchwald coupling to install secondary amine functionality, iix) N-phenyl triflimide; then, Pd$_2$(dba)$_3$, *rac*-BINAP, Cs$_2$CO$_3$, sarcosine methyl ester, toluene (12% for two steps). Dashed arrows indicate reactions yet to be run.

compound **19**.[35] Following standard procedures,[32] the nitro coumarin should be readily accessible, which can then be converted into the functionalized azidocoumarin **16**.

In our studies towards these functionalized 3-azidocoumarins, we initially pursued installation of the tertiary amine, in one step, from triflated 7-hydroxycoumarin via a palladium-mediated Buchwald coupling.[36] Transformations on the model system were performed with N-methyl sarcosine. While product was detected, yields were quite low, and the decision was made to follow the above route. It should be noted that Buchwald reactions were not optimized for this study, and future attention may be paid to synthetic routes utilizing this reaction.

Thus, significant progress has been made towards NCP-mediated dimerizations of hairpin polyamides, as well as towards a system in which dimerization can be monitored by fluorescence. It should be noted that for NCP-mediated dimerizations involving 6-

ring hairpin polyamides, compounds **1–15** are complementary to the designed supergroove sites and can be used without further synthesis.

Discussion

Initial experiments were designed to determine whether duplex DNA could template the dimerizations of hairpin polyamides in a turn-to-turn orientation. A panel of thermal, water-tolerant reactions was screened against duplex DNA templates with binding sites ranging from 10 to 13 base pairs. In order for templated reactions to be useful in *in vivo* experiments, reactive partners must not react in solution, but must form significant product in a biologically relevant time scale. On duplex DNA, only the Hüisgen 1,3-dipolar cycloaddition between an ethyl azide and a terminal alkynyl amidate (**3 + 5**) fits the above criteria.

The cycloaddition between **3** and **5** shows excellent efficiency on the 10 base pair template, achieving almost 50% yield in eight hours, and eventually reaching 78% completion in 24 hours. This represents an almost 25,000-fold increase in rate from the nontemplated reaction. While the turn-to-turn templated reaction is more efficient than its turn-to-tail counterpart,[20] it exhibits reduced site-separation specificity, reacting to modest yields on the 11 base pair site as well.

It was hypothesized that for the templated turn-to-tail dimer forming reactions[20] at 1 μM, product formation on the mismatch template was due to the ability of the polyamides to bind their mismatch sites at these high concentrations. When the concentration is decreased, the polyamides are better able to selectively bind the match site, and product formation on the mismatch templates is reduced. While the site

separation specificity is not great for the turn-to-turn dimerization reaction, the template mismatch specificity is exquisite. At 500 nM concentrations, the reaction proceeds on the match template more than an order of magnitude (11-fold) more efficiently than on the mismatch template.

With respect to the 1,3-dipolar cycloaddition between **3** and the methyl alkyne **6**, it is interesting that no reaction is observed on the 11 base pair template, even though the reactive species are identical in length to **3** and **5**. Perhaps on the longer 11 base pair template, the reactive groups are fully stretched out, and the transition state pushes the terminal methyl group into one of the walls or the floor of the minor groove, whereas at the shorter 10 base pair template, the linker is able to fold slightly to avoid such a steric clash.

The 1,3-dipolar cycloadditions between the bulky phenyl azide and the various alkynes did not proceed on duplex templates. Previous researchers have found that phenyl azides are more reactive than alkyl azides in these cycloadditions.[37, 38] Thus, the lack of reactivity must be due to some steric constraint imposed upon the reaction by sequestration in the minor groove. This reaction offers great potential for a supergroove-specific, turn-to-turn, dimer-forming reaction.

With respect to the other reactions tried, several interesting features emerge. First, the Wittig reaction does not proceed on any of the templates. The Wittig starting materials each contain bulky phenyl rings, with polyamide **15** containing the extremely bulky triphenylphosphine moiety. It is not surprising then, that this pairing produces no reaction within the narrow confines of the minor groove of the duplex templates. Perhaps

when the reaction is tried across the more sterically free nucleosomal supergroove, product will be observed.

Similarly, the Diels-Alder reactions result in bulky fused ring systems that may be incompatible within the narrow confines of the minor groove. That no templated reaction was observed may be due to a sterically large transition state that can not fit within the minor groove. Because the Diels-Alder reactions produce no product under non-templated conditions, these may be ideal reactions for dimer formation on nucleosomal supergroove templates that will not impose as stringent steric constraints on the transition state as duplex template.

With respect to the Michael additions, it is unfortunate that the reaction proceeds at 1 µM concentrations in the absence of template. Because of the significant non-templated product formation, this reaction cannot be used for dimerization across "supergroove" templates. It does appear that templates **A** and **B** were able to increase the yield (and by extension, rate) of product formation, however, the non-templated background rate is too high to warrant its use for biological applications. It is interesting that on the longer templates, these reactions do not produce any product. Perhaps the polyamides are bound to the minor groove, effectively sequestering the reactive groups in a geometry that does not allow the reactive species to interact, thereby inhibiting product formation.

Conclusion.

Linear, duplex DNA is able to template the formation of turn-to-turn hairpin dimers. While a panel of reactions was examined, only the 1,3-dipolar cycloaddition

between an alkyl azide and an alkynyl amidate proceeded in good yields and with good site-size and mismatch selectivity. At micromolar concentrations, duplex template enhances the rate of a dipolar cycloaddition between **3** and **5** 20,000-fold. Because turn-functionalized polyamides have been shown to have increased nuclear uptake ability relative to their tail-functionalized counterparts, this scaffold offers the opportunity to be used for the *in vivo* creation of hairpin dimers, potentially using the promoter of interest to guide formation of a molecule able to influence cell function.

Materials and Methods.

3-azido propionic acid (3a)[39]

Sodium azide (897 mg, 13.8 mmol) was dissolved in 5 mL water at 10 ºC. Keeping the temperature under 25 ºC, β-propiolactone (1 g, 13.8 mmol) was added portion-wise. The reaction was stirred at rt for 4 h, after which 8 mL of 37% HCl was added. The product was extracted with diethyl ether (3 x 20 mL). The combined extracts were dried over Na_2SO_4, and concentrated by rotary evaporation under reduced pressure. Spectra as reported in Leffeler, et al.[39]

Polyamide 3

Polyamide **1** (2 μmol) was combined with **3a** (3 μmol), PyBOP (3 μmol), and DIEA (20 μmol) in 300 μL anhydrous DMF at rt for 1 h. Reaction was quenched by addition of 9 mL 0.1% TFA in water, and purified by reversed phase HPLC for a final yield of 45%. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1020.1. Found 1019.6.

Polyamide 4

Synthesized from polyamide **1** (1 μmol) and the p-azido benzoic acid NHS-ester (Pierce, 2 μmol) in 200 μL 10:1 DMF:DIEA. Reaction was quenched by addition of 9 mL 0.1% TFA in water, and purified by reversed phase HPLC for a final yield of 72%. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1067.5. Found 1067.8.

Polyamide 5

For preparation, see Chapter 2 of this thesis.

Polyamide 6

Synthesized as **3** from 2-butynoic acid for an overall yield of 35%. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1060.5. Found 1060.8.

Polyamide 7

For preparation, see Chapter 2 of this thesis.

Polyamide 8

Synthesized as **4** from the NHS ester of γ-maleimido butyric acid (Pierce) for a 70% overall yield. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1088.3. Found 1088.3.

Polyamide 9

Synthesized as **3** from 2-furyl propionic acid for a 64% overall yield. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1141.5. Found 1141.8.

Polyamide 10

Synthesized as **3** from Boc-β-alanine, followed by deprotection with 50% TFA:DCM for a 41% overall yield. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1066.1. Found 1066.6.

Polyamide 11

Synthesized as **4** from the NHS ester of 1-S-acetyl acetic acid (Pierce), followed by deprotection with 2 M NaOH in 21% overall yield. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1068.5. Found 1068.5.

Polyamide 12

Synthesized as **4** from 4-maleimido cyclohexyl-1-carboxylic acid (Acros Organics) for a 71% overall yield. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1044.5. Found 1044.7.

Polyamide 13

Synthesized as **12** for a 71% overall yield. UV (H_2O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1116.5. Found 1116.5.

Polyamide 14

Synthesized as **4** from Salicaldehyde (Aldrich) for a 54% overall yield. UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1054.5. Found 1054.7.

Polyamide 15

Polyamide **2** was combined with the NHS ester of 4-iodoaminoacetic acid (SIAB, Pierce), according to **4**. After HPLC purification and lyophylization, product was mixed with 4-(diphenylphosphino) benzoic acid in 3:7 MeOH:THF for 8 h at 55 ºC for a two step yield of 12% . UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1459.2. Found 1459.6.

17

2-aminophenol (5 g, 47.6 mmol) was refluxed in ethyl formate (75 mL) for 48 h. The reaction was concentrated by rotary evaporation, and the product purified by silica gel chromatography (1:1 EtOAc:Hexanes) (3 g, 25 mmol, 51 % yield). The resulting purified product (540 mg, 3.96 mmol) was dissolved in 30 mL dry THF and cooled to 0 ºC. 11.3 mL of a 1 M solution of BH$_3$-THF was added via syringe. The reaction was stirred at 0 ºC for 1 h and then rt for 1.5 h. The reaction was quenched by the addition of 8 mL 10% citric acid. The product was extracted with EtOAc (5 x 20 mL). The combined organic layer was washed with brine (2 x 30 mL), dried over anhydrous Na$_2$SO$_4$, and concentrated by rotary evaporation. Product was purified by silica gel

chromatography (430 mg, 86% yield). ^{1}H NMR in CDCl$_3$ δ 7.0 m 1H, 6.2 m 2H, 6.08 m 1H, 2.79 s 3H, 2.05 s 1H.

18

Compound **17** (100 mg, 0.812 mmol) was dissolved in DMF (3 mL). 2,6 lutidene (600 μL) and methylbromoacetate (150 μL, 0.934 mmol) were added and the reaction stirred at rt for 20 h. The reaction was concentrated by rotary evaporation and the product purified by silica gel chromatography (2:1 Hexanes:EtOAc). Product was isolated as a white powder (163 mg, o.69 mmol, 85%). ^{1}H NMR in CDCl$_3$ δ 7.0 t 1H, 6.15 m 3H, 6.0 s, 1H, 3.82 s 2H, 3.21 s 3H, 1.56 s 3H.

19

Compound **18** (160 mg, 0.675 mmol) was dissolved in 1 mL DMF. POCl$_3$ (103 mg, 0.675 mmol) and DMF (492 mg, 6.75 mmol) were combined in a separate flask and stirred under N$_2$ at 0 ºC for 1 h and then rt for 1.5 h. This mix was then cannulated into the stirring solution of **18**. The reaction was then stirred at rt for 48 h. The reaction was quenched with 10 mL of 0.5 M NaOH. The pH was then adjusted to 7.0 and the product extracted into EtOAc. The combined organic layer was washed with brine, dried over sodium sulfate and concentrated by rotary evaporation. The product was isolated as a clear oil after silica gel chromatography (2:1 Hexanes:EtOAc, 60 mg, 35%). ^{1}H NMR in CDCl$_3$ δ 11.3 s 1H, 9.28 s 1H, 7.14 d 1H, 6.0 d 1H, 5.9 dd 1H, 3.95 s 2H, 3.31 s 3H, 1.62 s 3H.

Template-Derived Masses for Dimer Products.

3 + 5 - MALDI-TOF-MS calcd. (M + H): 2067.7. Found 2067.6.

3 + 6 - MALDI-TOF-MS calcd. (M + H): 2080.9. Found 2081.4.

3 + 7 - MALDI-TOF-MS calcd. (M + H): 2093.9. Found 2093.6.

8 + 10 - MALDI-TOF-MS calcd. (M + H): 2155.2. Found 2156.0.

8 + 11 - MALDI-TOF-MS calcd. (M + H): 2152.4.   Found 2178.2 [M+H+19]$^+$ (corresponds to hydrolyzed maleimide).

Sample Buchwald Coupling Procedures (20)

7-hydroxycoumarin (500 mg, 3.08 mmol) was combined in THF (20 mL) with triethylamine (429 μL, 3.08 mmol) and stirred for 5 min. N-phenyltriflimide (1.1 g, 3.08 mmol) was then added and the reaction stirred under inert atmosphere for 3 h at rt. The reaction was concentrated and the residue purified by silica gel chromatography (3:1 Hexanes:EtOAc). Product was isolated as a white powder (760 mg, 83%). In a flame-dried flask, the triflated material (100 mg, 0.34 mmol), Cs_2CO_3 (220 mg, 0.68 mmol), (±)-BINAP (32 mg, 0.051 mmol), $Pd_2(dba)_3$ (35 mg, 0.034 mmol), and N-methyl glycine methyl ester (57 mg, 0.41 mmol) were combined and dried under vacuum for 2 h. Toluene (3 mL) was then added, and the mixture stirred at 80 ºC for 14 h. The reaction mixture was filtered, and the ppt. washed with DCM. The combined filtrates were concentrated by rotary evaporation, and the product chromatographed in 1.5:1 Hexanes:EtOAc. The product (**20**) was isolated as a bright yellow oil (10 mg, 12% yield).

Preparation of duplex DNA for kinetic experiments. Duplex DNA was prepared by incubating equal amounts of complementary sets of synthetic oligonucleotides at 90 °C for 10 min, then slowly allowing them to cool to rt. Resulting duplex DNA was quantified by UV by the relationship 1 OD_{260} unit = 50 μg/mL duplex DNA. Duplex DNA was stored at -20 °C in water.

Typical Reaction Procedure. All kinetic reactions were performed in 1.7 mL presiliconized microcentrifuge tubes obtained from VWR International. Total reaction volumes were 1.2 mL aqueous solutions of equamolar concentrations of each hairpin polyamide and DNA (2 mM Tris-HCl, 2 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.0, 37 °C). Reactions were monitored by HPLC by direct injection of reaction samples onto a RAINEN C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed-phase column in 0.1% (w/v) TFA-H_2O with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 0.5% CH_3CN/min. Peaks were quantified using the Beckman Coulter GOLD software package. Verification of product was determined by MALDI TOF-MS of ~40 μL samples of each reaction concentrated on a ZipTip 2 mg C_{18} pipette tip eluted with 75% (v/v) acetonitrile in 0.1% (w/v) TFA.

References.

1. Dervan, P. B., *Science*, **1986,** *232*, 464–471.

2. Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2001,** *9*, 2215–2235.

3. Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology*, **2003,** *13*, 284–299.

4. Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2004,** *101*, 16768–16773.

5. Livengood, J. A.; Fechter, E. J.; Dervan, P. B.; Nyborg, J. K., *Frontiers in Bioscience*, **2004,** *9*, 3058–3067.

6. Nguyen-Hackley, D. H.; Ramm, E.; Taylor, C. M.; Joung, J. K.; Dervan, P. B.; Pabo, C. O., *Biochemistry*, **2004,** *43*, 3880–3890.

7. Fechter, E. J.; Dervan, P. B., *Journal of the American Chemical Society*, **2003,** *125*, 8476–8485.

8. Ehley, J. A.; Melander, C.; Herman, D.; Baird, E. E.; Ferguson, H. A.; Goodrich, J. A.; Dervan, P. B.; Gottesfeld, J. M., *Molecular and Cellular Biology*, **2002,** *22*, 1723–1733.

9. Bremer, R. E.; Wurtz, N. R.; Szewczyk, J. W.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2001,** *9*, 2093–2103.

10. Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A., *Journal of Biological Chemistry*, **2000,** *275*, 24246–24254.

11. Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Graves, B. J.; Gottesfeld, J. M., *Journal of Biological Chemistry*, **1999,** *274*, 12765–12773.

12. Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B., *Proceedings of The National Academy of Sciences of the United States of America*, **2003,** *100*, 12063–12068.

13. Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B., *Nucleic Acids Research*, **2004,** *32*, 2802–2818.

14. Bailly, C.; Chaires, J. B., *Bioconjugate Chemistry*, **1998,** *9*, 513–538.

15. Faria, M.; Giovannangeli, C., *Journal of Gene Medicine*, **2001,** *3*, 299–310.

16. Reddy, B. S. P.; Sharma, S. K.; Lown, J. W., *Current Medicinal Chemistry*, **2001,** *8*, 475–508.

17. Herman, D. M.; Baird, E. E.; Dervan, P. B., *Chemistry-a European Journal*, **1999,** *5*, 975–983.

18. Kers, I.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 3339–3349.

19. Weyermann, P.; Dervan, P. B., *Journal of the American Chemical Society*, **2002,** *124*, 6872–6878.

20. Poulin-Kerstien, A. T.; Dervan, P. B., *Journal of the American Chemical Society*, **2003,** *125*, 15811–15821.

21. Edayathumangalam, R. S.; Weyermann, P.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Proceedings of the National Academy of Sciences of the United States of America*, **2004,** *101*, 6864–6869.

22. Poulin-Kerstien, A. T. DNA-Templated Dimerizations of Minor Groove-Binding Polyamides. California Institute of Technology, Pasadena, 2005.

23. Calderone, C. T.; Puckett, J. W.; Gartner, Z. J.; Liu, D. R., *Angewandte Chemie-International Edition*, **2002,** *41*, 4104–4108.

24. Gartner, Z. J.; Kanan, M. W.; Liu, D. R., *Angewandte Chemie-International Edition*, **2002,** *41*, 1796–1798.

25. deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society*, **1997,** *119*, 7909–7916.

26. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nature Structural Biology*, **1998,** *5*, 104–109.

27. 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.

28. Suto, R. K.; Edayathumangalam, R. S.; White, C. L.; Melander, C.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Journal of Molecular Biology*, **2003,** *326*, 371–380.

29. Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6141–6146.

30. Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 2767–2774.

31. Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J., *Journal of Molecular Biology*, **2002,** *319*, 1097–1113.

32. Sivakumar, K.; Xie, F.; Cash, B. M.; Long, S.; Barnhill, H. N.; Wang, Q., *Organic Letters*, **2004,** *6*, 4603–4606.

33. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., *Angewandte Chemie-International Edition*, **2002,** *41*, 2596–2599.

34. Nakagawa, Y.; Irie, K.; Masuda, A.; Ohigashi, H., *Tetrahedron*, **2002,** *58*, 2101–2115.

35. Feuster, E. K.; Glass, T. E., *Journal of the American Chemical Society*, **2003,** *125*, 16174–16175.

36. Muci, A. R.; Buchwald, S. L., Practical palladium catalysts for C-N and C-O bond formation. In *Cross-Coupling Reactions*, 2002; Vol. 219, pp 131–209.

37. Huisgen, R., *1,3-Dipolar Cycloaddition Chemistry*. Wiley-Interscience: New York, 1984; p Chapter 1.

38. Huisgen, R., *Profiles, Pathways, and Dreams*. American Chemical Society: Washington, DC, 1994.

39. Leffler, J. E.; Temple, R. D., *Journal of the American Chemical Society*, **1967,** *89*, 5235–5246.

Chapter 4

Controlling Polyamide Dimerization through Non-Covalent Interactions

Abstract

Due to size limitations on the cellular uptake of polyamides, we herein investigate whether side-by-side dimerization of polyamides can be accomplished through non-covalent interactions. A series of novel pyrrole rings were synthesized and incorporated into 5-ring polyamides. We find that functionalization of the N1 position of pyrrole with positively interacting side chains is unable to control polyamide dimerization. Functionalization with negatively interacting side chains is able to control hetero- and homodimerization.

Introduction

While increasing the number or pyrrole and imidazole rings in a given polyamide increases binding affinity, cellular uptake studies have shown that polyamides containing more than eight rings are unable to reach their target DNA within the nuclei of live cells. The goal of the research reported herein is to design polyamide systems capable of targeting large (6–10 bp) sequences of DNA with high affinity and specificity, while at the same time possessing favorable cellular and nuclear uptake properties.

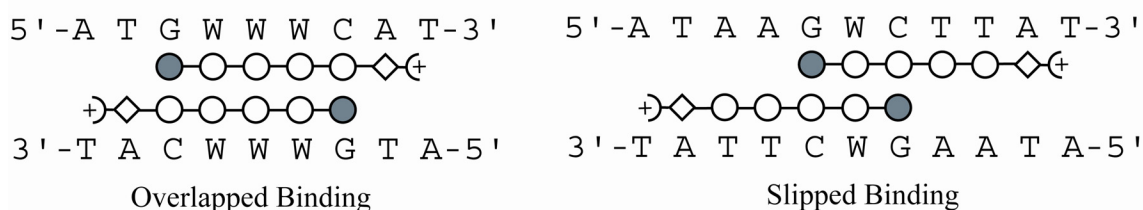Nature uses a variety of non-covalent forces to create specific protein folds and to associate multiple proteins into complexes.[1] A notable example is the ribosomal complex, which consists of over 30 proteins and ribonucleic (RNA) fragments held together by non-covalent interactions.[2] These forces include van der Waals forces, hydrogen bonding, Coulombic forces, and cation-π interactions.[1, 3-6] Non-covalent forces are likewise responsible for the binding of polyamide molecules to DNA. While linking the two strands of the 2:1 polyamide/DNA complex covalently greatly increases the binding affinity, the size of the polyamide is doubled and cellular uptake may be compromised. Perhaps unlinked polyamide strands can be functionalized in such a way that imparts some non-covalent associative force between the two strands, leading to increased affinity and specificity. However, because the forces are non-covalent, and the polyamide strands are not associated until bound to DNA, the smaller polyamide strands may possess favorable uptake properties. Thus, with this system, both the goal of cellular uptake and that of specific and high affinity recognition of DNA may be realized.

2:1 vs 1:1 Polyamide: DNA Stoichiometry

```
5'-A T G W W W W A T-3'
      ●─○─○─○─○─◇─(
3'-T A C W W W W T A-5'
        1:1 Binding
```

```
5'-A T G W W W C A T-3'
      ●─○─○─○─○─◇─(
   )─◇─○─○─○─○─●
  3'-T A C W W W G T A-5'
          2:1 Binding
```
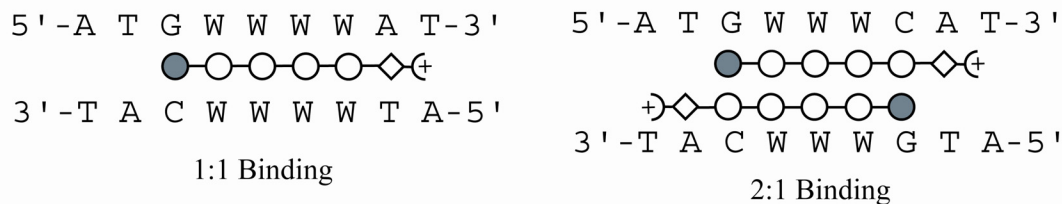
Fully Overlapped vs. Slipped

```
5'-A T G W W C A T-3'
      ●─○─○─○─○─◇─(
   )─◇─○─○─○─○─●
3'-T A C W W W G T A-5'
    Overlapped Binding
```

```
5'-A T A A G W C T T A T-3'
          ●─○─○─○─○─◇─(
   )─◇─○─○─○─○─●
3'-T A T T C W G A A T A-5'
        Slipped Binding
```

Distinct Fully Overlapped Sites

```
5'-A T G W W C A T-3'
      ●─○─○─○─○─◇─(
  )─◇─○─○─○─●
3'-T A C W W W G T A-5'
   1 imidazole homodimer
```

```
5'-A T G G W C C A T-3'
      ●─●─○─○─○─◇─(
  )─◇─○─○─○─●─●
3'-T A C C W G G T A-5'
    2 imidazole homodimer
```

```
5'-A T G G W W C A T-3'
      ●─●─○─○─○─◇─(
   )─◇─○─○─○─○─●
3'-T A C C W W G T A-5'
      Heterodimer Binding
```
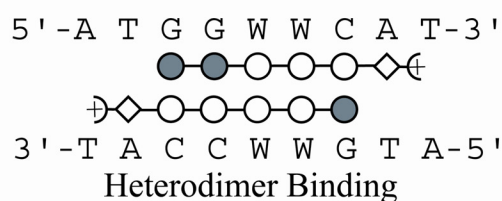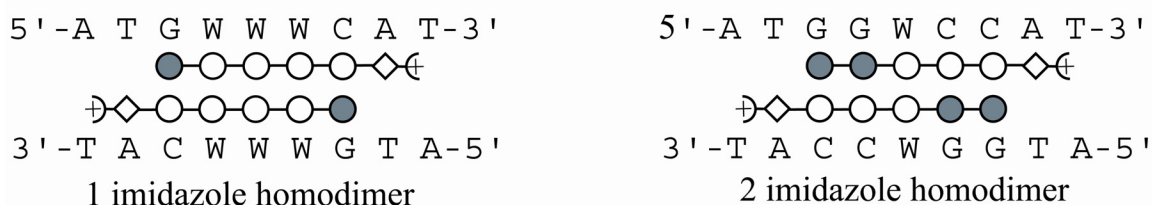
Figure 4.1. Schematic model for illustrating the various binding modes possible when DNA is incubated with two different polyamides. Imidazole and pyrrole are represented by dark and light circles, respectively.

Dimeric DNA/polyamide complexes are inherently ambiguous. As shown schematically in Figure 4.1, when DNA is incubated with two different polyamide strands, multiple binding sites can be targeted. New functional groups incorporated onto unlinked dimers must be able to select for one or a few of the myriad binding modes while at the same time not interfering with the polyamide's DNA recognition elements.

The problem of heterodimer versus homodimer formation within the fully overlapped, 2:1 binding motif is the subject of this research report.

Polyamide Functionalization.

When polyamides bind to DNA the N1-methyl position of each pyrrole and imidazole subunit points up from the floor of the minor groove into solvent. Additionally, paired pyrrole and imidazole residues place their N-methyl groups in close proximity.[7] Functionalization of polyamides at this position would be ideal because groups placed here would be able to interact based on their close proximity, and would not interfere with the DNA-recognition side of the polyamide (Figure 4.2).
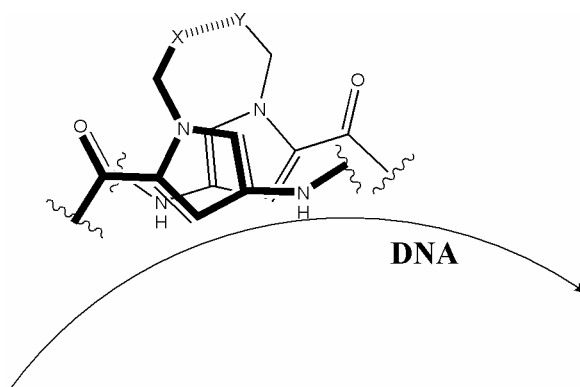
Figure 4.2. Structure of two polyamide strands bound dimerically in the minor groove of DNA. Functionalization of the polyamides with "X" and "Y" at the N1 position of pyrrole will place new moieties in close proximity without interfering with the DNA-recognition face of the polyamide.

The polyamide ImPyPyPyPy-β-Dp (**PA1**) had previously been analyzed by DNase I footprinting and was shown to bind with submicromolar affinity.[8] In this study, **PA1** is tested along with the polyamide ImImPyPyPy-β-Dp (**PA2**). The designed plasmid incorporates binding sites for both **PA1** and **PA2** homodimers as well as the site for the

PA1 = ●–○–○–○–○–◇–⊹   PA2 = ●–●–○–○–○–◇–⊹

```
5'-G GATC CTTAA GTTCGTGGGC CATGGTACAT TCGTGGGCCA TGGTCCATTC GTGGGCCATG TTACATTCGA AGCT TGGCGT-3'
3'-C CTAG GAATT CAAGCACCCG GTACCATGTA AGCACCCGGT ACCAGGTAAG CACCCGGTAC AATGTAAGCT TCGA ACCGCA-5'
              PA1/PA2              PA2/PA2              PA1/PA1
```
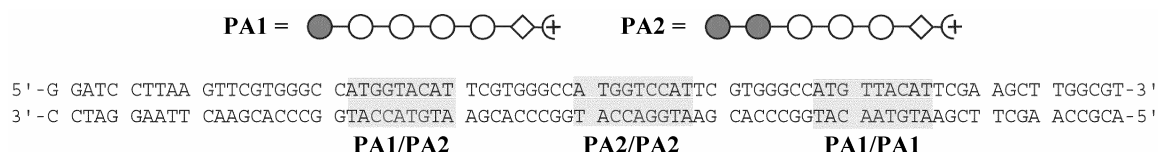
Figure 4.3. Sequence of plasmid insert pDHN2 used to probe homodimer vs. heterodimer formation. The plasmid contains binding sites for each of the two homodimers (center, right) and for the heterodimer (left).

PA1/PA2 heterodimer (Figure 4.3). This plasmid is designed to probe whether polyamide N1 functionalization will be able to control the relative affinities of polyamide dimers for the three distinct sites. This system allows for the central Py/Py pair to be functionalized as in Figure 4.2, and non-covalent interactions probed. Figure 4.4 shows schematically the functional groups probed in this study. Several biologically relevant interactions such as sterics, hydrogen bonding, quadrapole interaction, and cation-π are represented in this series. New pyrrole rings functionalized as shown below were synthesized, incorporated into polyamides, and their ability to drive specific dimer formation probed by quantitative DNase I footprinting.[9]

Synthesis of New Monomers (Figure 4.5).

A novel isobutyl imidazole trichloroketone was used as a replacement for the 1-methyl imidazole cap, was made in 84% yield from compound 2-trichloroketo imidazole, which was prepared according to published procedures.[10]

Monomer **7i** was prepared by Nick Wurtz. All other Boc-protected, N1-functionalized pyrrole amino acids were synthesized as shown in Figure 4.5. N1 alkylation was achieved from the 4-nitro or 4-NHBoc 2-ester pyrroles (**1**, **2**, **3**, **4**) using various alkyl bromides of the form Br-R (R = **a-f**, **h**, **k**), tetrabutylammonium iodide, and potassium carbonate in acetone with yields ranging from 70–99%. For the base-labile phthalimide-
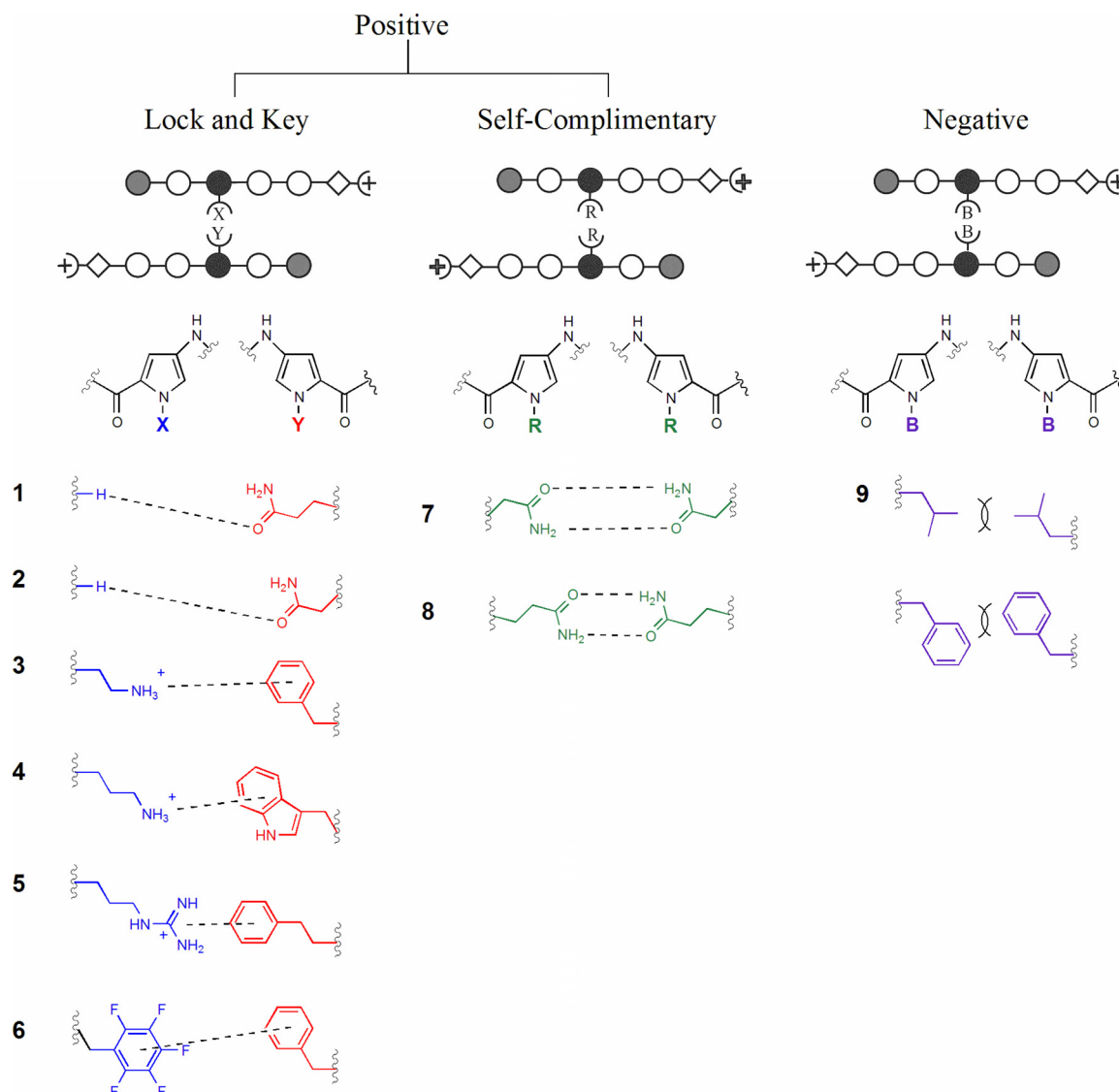
Figure 4.4. Schematic and structural representation of the non-covalent interactions reported herein. **Lock-and-Key** (left): Two different functional groups provide a stabilizing interaction. Examples include hydrogen bonding (1, 2), Cation-π (3-5), and quadrapole interactions (6). **Positive Interaction** (center): Two identical side chains interact to provide a stabilizing force. Examples include hydrogen bonding (7, 8) between two primary amides. **Negative Interaction** (right): Two identical side chains interact negatively to destabilize the pairing. Examples include steric exclusion from bulky groups (9, 10). Associative forces are represented by dashed lines; steric clash is represented by overlapping arcs.

protected amines (**a** and **e**), the pyrrole trimethylsilyl ethyl (TMSE) ester[11] was used in order to effect ester hydrolysis using tetrabutylammonium fluoride without disturbing the phthalimide moiety. The TMSE ester was also used when making pentafluorobenzyl derivative **b** in order to avoid degradation of the product during the basic hydrolysis of
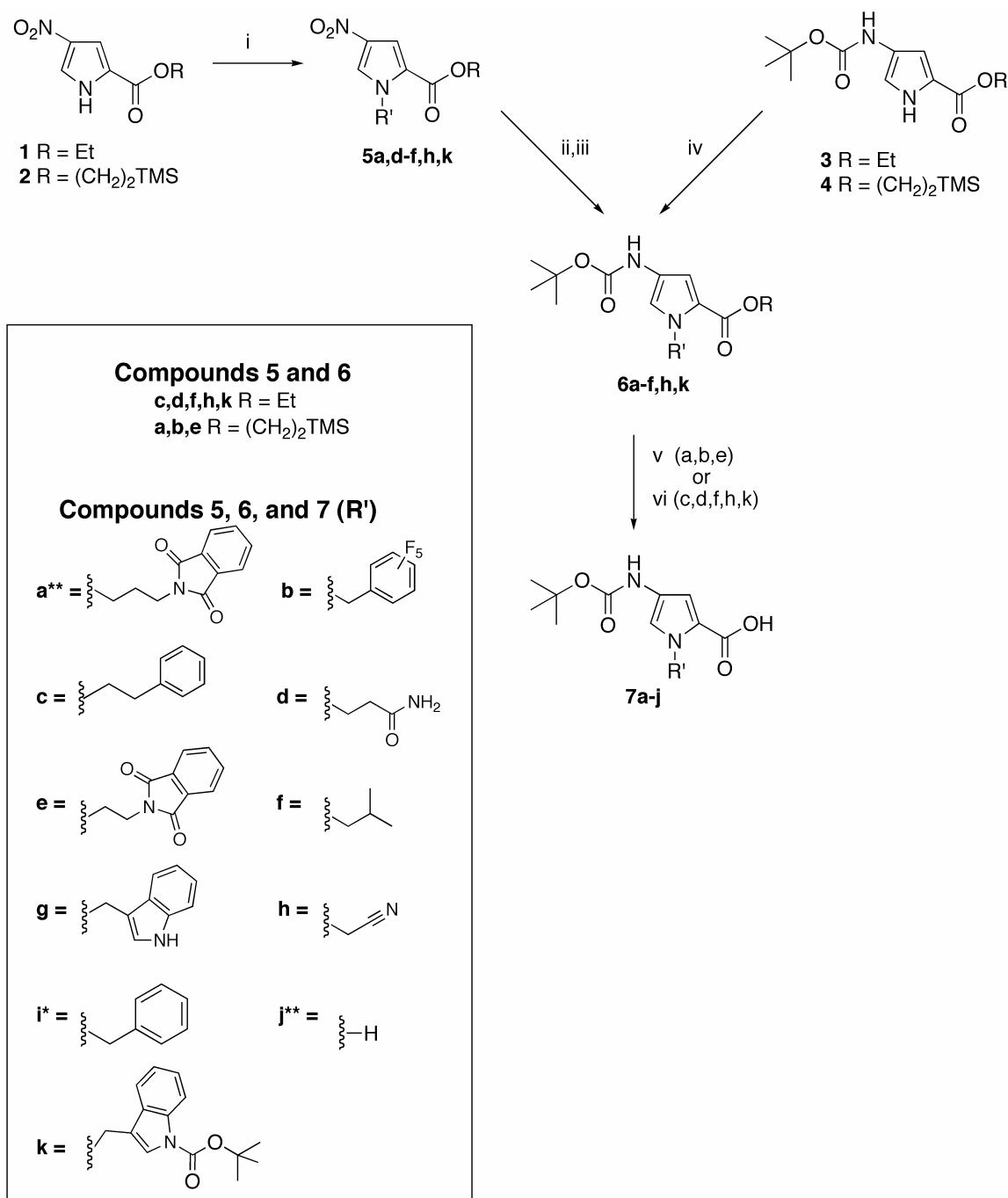
Figure 4.5. Synthetic scheme for novel N-functionalized pyrrole monomers. i) Br-R(**a**, **d–h**), K$_2$CO$_3$, Bu$_4$NI, acetone; ii)10% Pd/C, H$_2$, EtOAc; iii) Boc$_2$O, 1M NaHCO$_3$; iv) Br-R (**b**, **c**), K$_2$CO$_3$, Bu$_4$NI, acetone; v) TBAF, THF; vi) 2M NaOH, EtOH. (*) Compound provided by Nick Wurtz. (**) Compound prepared as previously published.

the ethyl ester precursor. Alkylations with both the pentafluorobenzyl and ethylbenzyl

bromides were performed on the more advanced 4-NHBoc 2-TMSE ester pyrrole (**4**)

because of the greater reactivity of the bromides. All other N1- modifications were carried out via alkylation of the nitro ethyl ester pyrrole **1**.[12] Subsequent reduction using palladium on carbon under an atmosphere of hydrogen gas, followed by Boc- protection using Boc anhydride, afforded **6a, d, e, f,** and **g** in high yields. The Boc-protected carboxylic acid final products **7a–j** were obtained in high yields from either the ethyl- or TMSE- protected intermediates **6a–f, h, k**. Treatment of **6k** with base hydrolyzed the ester and deprotected the Boc-indole in one step, affording **7g** in 97% yield.

Synthesis of Polyamides.

Unfunctionalized Polyamides (**PA1** and **PA2)** were synthesized from pyrrole and imidazole Boc-protected amino acids on solid support.[13] Polyamides **PA3–PA5**, **PA8–PA15**, and **PA17–PA24** (Figure 4.6) were synthesized on solid support using monomers **7a–j** and **9** along with the standard imidazole and pyrrole monomers (Figure 4.7). Coupling of the isobutyl trichloroketone cap (**9**) was carried out using 1.0 equivalent of resin-bound amine, 2.0 equivalents of **9**, and 1.0 equivalents of diisopropylethylamine (DIEA) in DMF at 37 °C for two hours in high conversion as determined by reversed-phase HPLC. Polyamide ImPyPy(3G)PyPy-β-Dp (Propylguanidyl side chain:**PA16**) was synthesized in 5% isolated yield from resin-bound ImPyPy(3P)PyPy-β-RESIN (Propylamino side chain) and N,N'-Boc-guanidyl pyrazole followed by deprotection. Initial attempts to synthesize polyamides ImPy(Am)PyPyPy-β-Dp (amide side chain: **PA6**) and ImImPyPy(Am)Py-β-Dp (amide side chain: **PA7**) were carried out using a Boc-protected pyrrole carboxylic acid functionalized on N1 with the primary amide

group (-CH$_2$CONH$_2$). However, activation of this monomer using standard activating

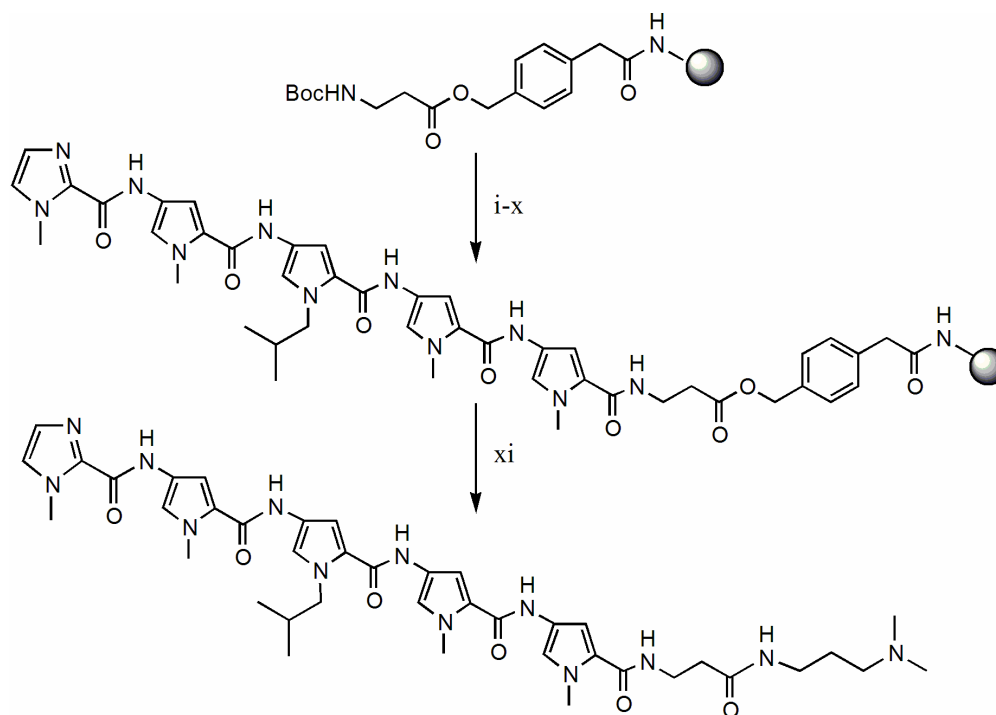agents (DCC/HOBt, HBTU, Py-BOP) led to the intramolecular attack of the amide

Figure 4.6. Representative solid-phase synthetic scheme for ImPyPy(iBu)PyPy-β-Dp (**PA3**) from Boc-β-Ala-PAM resin. i) 50% TFA:DCM; ii) Boc-Py-OBt, DIEA, NMP; iii) 50% TFA:DCM; iv) Boc-Py-OBt, DIEA, NMP; v) 50% TFA:DCM; vi) Boc-Py(iBu)-OH, HBTU, DIEA, NMP; vii) 50% TFA:DCM; iix) Boc-Py-OBt, DIEA, NMP; ix) 50% TFA:DCM; x) Im-OH, HBTU, DIEA, NMP; xi) Dimethylamino propyl amine, 55 °C, 12 h.

functionality on the activated ester, forming an unreactive bicyclic by-product. To

overcome this difficulty, an attempt was made to form **PA6** and **PA7** by alkylating

ImDsPyPyPy-β-Dp (**PA10**) and ImImPyDsPy-β-Dp (**PA11**) (Ds = Des-methyl pyrrole

monomer[12]) with potassium carbonate and bromoacetamide. ^{1}H NMR analysis of the

alkylated product showed that the tail dimethylammonium nitrogen was the site of

alkylation. Subsequent analysis shows that potassium carbonate is not basic enough to

deprotonate the pyrrole ring (pK$_a$ of pyrrole in DMSO = 23), thus making the

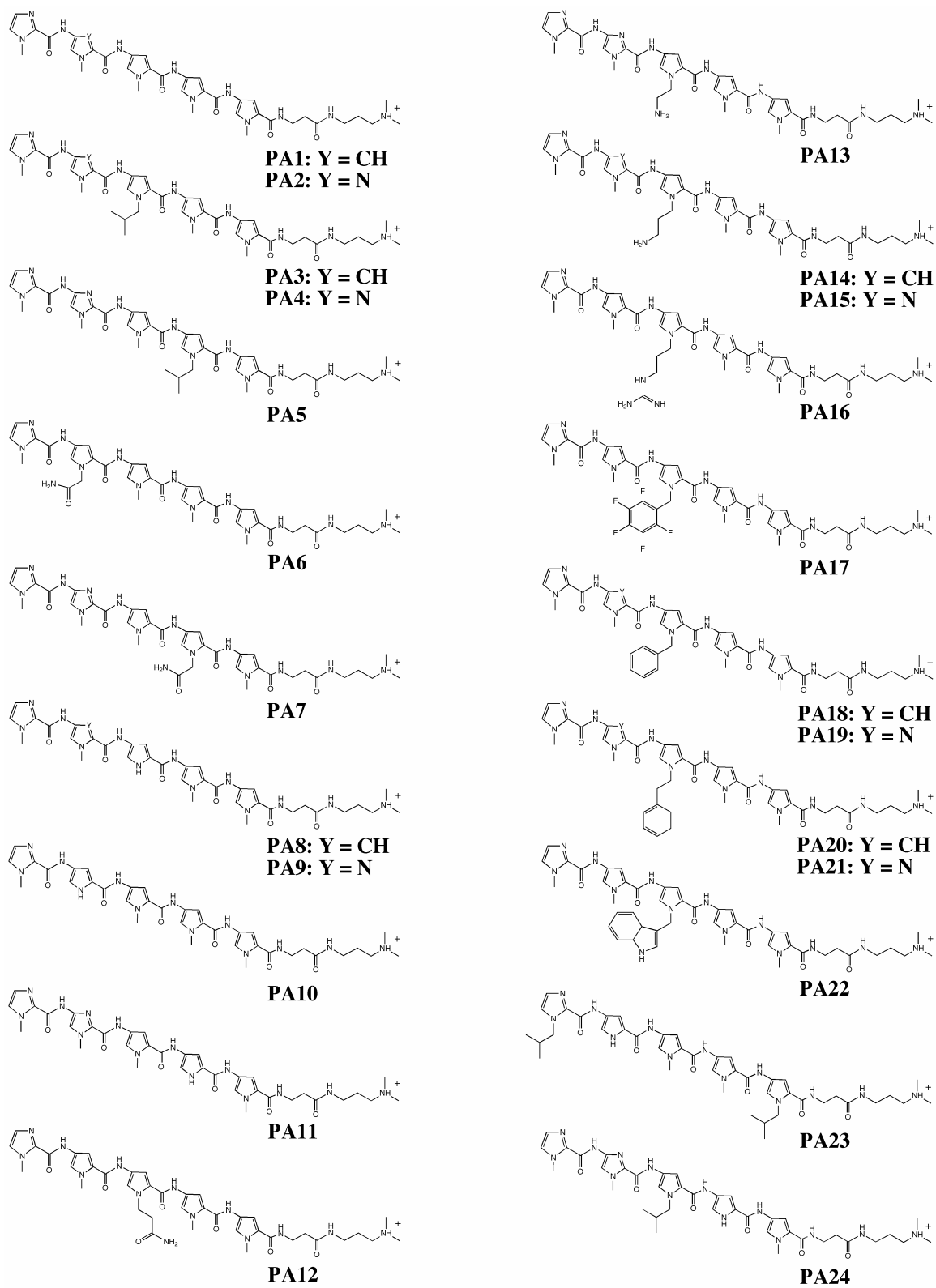dimethylalkylammonium nitrogen the most reactive nucleophile.

Figure 4.7. Chemical structures of the 5-ring polyamides synthesized on solid support.

By pre-incubating **PA10** and **PA11** with 10.0 equivalents of sodium hydride in DMF, a global deprotonation was achieved. Upon addition of one equivalent of bromoacetamide, the monoalkylated products **PA6** and **PA7** were obtained in 30% yield. Alternatively, monomer **7h** could be used in the standard solid-phase protocol. Upon acidic workup, the nitrile is hydrated to give **PA6** and **PA7** in 5% isolated yield. While the yields of the two methods seem very different, the 5% yield is the amount of isolated product after two steps, cleavage from resin and HPLC purification. The seemingly higher 30% yield represents the amount of product isolated after a single HPLC purification only. With respect to overall yield, the two methods do not differ significantly. The purity and identity of each polyamide was determined by reversed-phase HPLC, MALDI-TOF mass spectrometry, and ^{1}H NMR.

Quantitative DNase I Footprinting.

Equilibrium association constants (K_a) were determined using DNase I footprint titration experiments (10 mM Tris HCl, 10 mM KCl, 10 mM MgCl$_2$, and 5 mM CaCl$_2$, pH 7.0, 22 °C). Plasmid pDHN2[14] contains three match sites: 5'-TGTTACA-3' is a match site for the one-imidazole homodimer (**PA1/PA1**), 5'-TGGTCCA-3' is a match site for the two-imidazole homodimer (**PA2/PA2**), and 5'-TGGTACA-3' is a match site for the heterodimer formed between the one- and two-imidazole compounds (**PA1/PA2**). The parent polyamides were footprinted against pDHN2 and their binding affinities determined for reference. **PA1** binds as a homodimer to its match site with an affinity of 5.2×10^8 M^{-1}. **PA2** binds to its homodimer match site with an affinity of 4.8×10^8 M^{-1}. The **PA1/PA2** heterodimer binds its match site with a similar affinity of 4.2×10^8 M^{-1}

Figure 4.8. Quantitative DNase footprinting experiment with ImPyPyPyPy-β-Dp (**PA1**) and ImImPyPyPy- β -Dp (**PA2**) on the 3'-^{32}P-labeled restriction fragment pDHN2. Gel lanes from left to right: A reaction, G reaction, intact DNA, DNase I standard, DNase digestion products in the presence of 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM polyamide. Right: Quantitation and Hill plots of data from gel. Below each plot is the binding affinity constant obtained at each site (Ka in M^{-1}). Conditions and data analysis as outlined in the text.

(Figure 4.8). Association constants for functionalized polyamides are compared to these parent values to determine the functional group's ability to select for one of the three binding modes. As an initial screen for binding, compounds were tested on the one-imidazole homodimer match site with the side chains of interest paired across from each

other on the central pyrrole residue. Compounds that were either ambiguous or promising in this system were tested on the heterodimer site. Each equilibrium association constant is an average of at least three quantitative DNase I footprint titrations. The footprinting results are summarized in Table 4.1. With respect to the parent, unfunctionalized polyamides (**PA1** and **PA2**), several interesting features emerge from these data. Firstly, both ImPyPy(Bz)PyPy-β-Dp (benzyl side chain: **PA18**) and ImPyPy(iBu)PyPy-β-Dp (isobutyl side chain: **PA3**) exhibit a 16-fold reduction in binding affinity when homodimerically paired within the minor groove. Interestingly, ImPyPy(5FBz)PyPy-β-Dp (pentafluorobenzyl moiety: **PA17**), which is isosteric to the benzyl group while carrying the opposite quadrapole, does not alter binding affinity relative to the parent compound.

Alternatively, each of the positively charged side chains exhibits an increased affinity when homodimerically paired. Both ImPyPy(3G)PyPy-β-Dp (propylguanidyl side chain: **PA16**) and ImPyPy(3P)PyPy-β-Dp (propylamino side chain: **PA14**) bind to their match sites with a K_a of 6.6 x 10^9 M^{-1}. This affinity represents a greater than 10-fold increase over the parent polyamides. ImImPy(2P)PyPy-β-Dp (ethylamino side chain: **PA13**) gives only a 3-fold binding affinity increase. Likewise, the hydrogen side chain of ImPyDsPyPy-β-Dp (**PA8**) exhibits a 4-fold increase in affinity.

When examining heterodimeric side chain pairings, each of the proposed cation/π interactions between the positively-charged side chains (**PA13–PA16**) paired with the three aromatic side chains (**PA18–PA22**) exhibit affinities within a factor of 3 from the parent polyamides. The propylamino moiety of **PA15**, paired with either the benzyl (**PA18**) or ethylbenzyl (**PA20**) side chains shows a slight increase in affinity while the
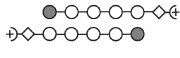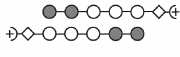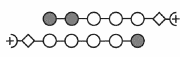
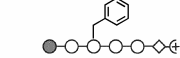**Table 2: Equilibrium Association Constants***
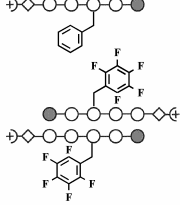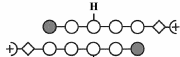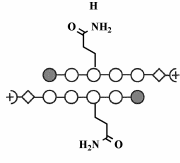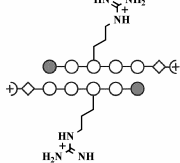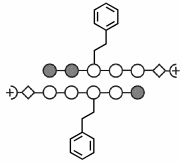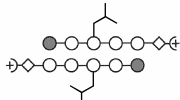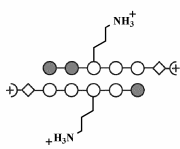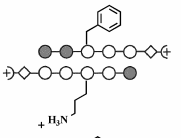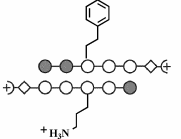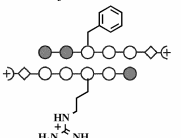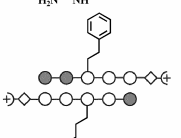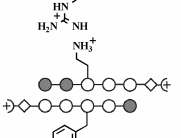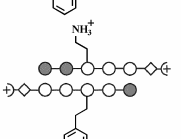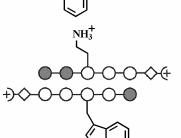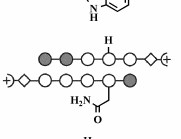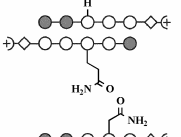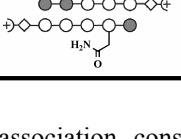
PA #	Polyamide	5'-TGGTACA-3'	5'-TGGTCCA-3'	5'-TGTTACA-3'	$\Delta Ka_{(parent)}$
1 + 1		$4.1 \times 10^7 \text{ M}^{-1}$	--------------	$5.1 \times 10^8 \text{ M}^{-1}$	1.0
2 + 2		$9.6 \times 10^7 \text{ M}^{-1}$	$5.4 \times 10^8 \text{ M}^{-1}$	--------------	1.0
2 + 1		$4.2 \times 10^8 \text{ M}^{-1}$	$4.6 \times 10^8 \text{ M}^{-1}$	$5.2 \times 10^8 \text{ M}^{-1}$	1.0
18 + 18		--------------	--------------	$2.7 \times 10^7 \text{ M}^{-1}$	.05
17 + 17		$5.1 \times 10^7 \text{ M}^{-1}$	--------------	$6.7 \times 10^8 \text{ M}^{-1}$	1.3
8 + 8		$1.2 \times 10^7 \text{ M}^{-1}$	--------------	$1.7 \times 10^9 \text{ M}^{-1}$	3.3
12 + 12		$8.5 \times 10^7 \text{ M}^{-1}$	--------------	$6.6 \times 10^8 \text{ M}^{-1}$	1.3
16 + 16		$4.0 \times 10^8 \text{ M}^{-1}$	$1.3 \times 10^8 \text{ M}^{-1}$	$6.5 \times 10^9 \text{ M}^{-1}$	11.0
21 + 20		$2.0 \times 10^7 \text{ M}^{-1}$	$2.0 \times 10^7 \text{ M}^{-1}$	$3.4 \times 10^7 \text{ M}^{-1}$	0.065
3 + 3		--------------	--------------	$2.8 \times 10^7 \text{ M}^{-1}$	0.05
15 + 14		$3.7 \times 10^9 \text{ M}^{-1}$	$2.0 \times 10^9 \text{ M}^{-1}$	$6.7 \times 10^9 \text{ M}^{-1}$	12.9

Table 2 (continued): Equilibrium Association Constants*

PA #	Polyamide	5'-TGGTACA-3'	5'-TGGTCCA-3'	5'-TGTTACA-3'	ΔK_a(parent)
19 + 14		1.1×10^9 M^{-1}	2.5×10^8 M^{-1}	5.7×10^9 M^{-1}	2.4
21 + 14		1.0×10^9 M^{-1}	2.6×10^8 M^{-1}	7.6×10^9 M^{-1}	2.4
19 + 16		3.5×10^8 M^{-1}	1.0×10^8 M^{-1}	2.1×10^9 M^{-1}	0.83
21 + 16		3.4×10^8 M^{-1}	9.8×10^7 M^{-1}	2.3×10^9 M^{-1}	0.83
13 + 18		2.1×10^8 M^{-1}	1.5×10^9 M^{-1}	1.8×10^8 M^{-1}	0.5
13 + 21		2.1×10^8 M^{-1}	1.5×10^9 M^{-1}	2.0×10^8 M^{-1}	0.5
13 + 22		9.8×10^7 M^{-1}	1.46×10^9 M^{-1}	2.4×10^8 M^{-1}	0.23
11 + 7		6.5×10^8 M^{-1}	9.5×10^8 M^{-1}	4.1×10^8 M^{-1}	1.5
11 + 12		1.2×10^9 M^{-1}	1.3×10^9 M^{-1}	6.6×10^9 M^{-1}	2.9
7 + 6		1.0×10^8 M^{-1}	1.8×10^8 M^{-1}	2.2×10^8 M^{-1}	0.23

*Equilibrium association constants are the mean values obtained from three quantitative DNase I footprinting experiments. The standard deviation for each data point is less that 10% of the value reported. Assays were carried out as reported in the text. Dashed lines indicated no binding at that site. Left column; Polyamide number (Figure 4.7) – First number corresponds to the top polyamide pictured in column 2. #: Where multiple match sites are available (heterodimer assays), the relative affinity given the ratio of binding affinities of the functionalized polyamides listed relative to the binding affinity of the unfunctionalized parent polyamides (**PA1 + PA2**) on the heterodimer site (5'-aTGGTACAt-3').

propylguanidyl (**PA16**) and ethylamino (**PA13**) side chains bind with slightly lower affinity when paired with the aromatic side chains.

The putative hydrogen-bonding interactions (**PA12** homodimer, **PA6** with **PA7**, **PA12** with **PA9**, and **PA10** with **PA7**) give binding affinities that were quite varied. With respect to the unfunctionalized polyamides, ImPyPy(2Am)PyPy-β-Dp (ethylamido side chain: **PA12**) homodimerically paired does not differ in affinity from the parent polyamides. In contrast, **PA12** paired with the ImImDsPyPy-β-Dp (**PA9**) gives a 3-fold increase in affinity over the unmodified system, binding with a K_a of 1.2 nM^{-1}. Both the primary amide (**PA6** (ImPy(Am)PyPyPy-β-Dp) with **PA7** (ImImPyPy(Am)Py-β-Dp)) homodimer and the primary amide (**PA7**) with ImDsPyPyPy-β-Dp (hydrogen side chain: **PA10**) heterodimer fail to increase polyamide affinity.

In order to assess the specificity of the isobutyl interaction, ImImPyPy(iBu)Py-β-Dp (**PA5**) was synthesized and footprinted with **PA3**. As shown in Figure 4.9, when the isobutyl side chains are paired against the standard methyl group, and separated from each other by one residue, over 70% of the affinity loss incurred by the isobutyl pairing is rescued. Furthermore, when the two isobutyl groups are separated by two residues, another 15% of the binding affinity is recovered.

Because the isobutyl functional group showed a pair-specific destabilizing effect, Im(iBu)DsPyPyPy(iBu)-β-Dp (**PA23**) and ImImPy(iBu)DsPy-β-Dp (**PA24**) were synthesized and probed for their ability to favor the heterodimeric binding motif over either of the two homodimeric binding motifs. Because each isobutyl group contributes a two-fold, non-specific reduction in affinity, the Des-methyl pyrrole monomer was incorporated in order to non-specifically boost the affinity of the polyamides. As shown
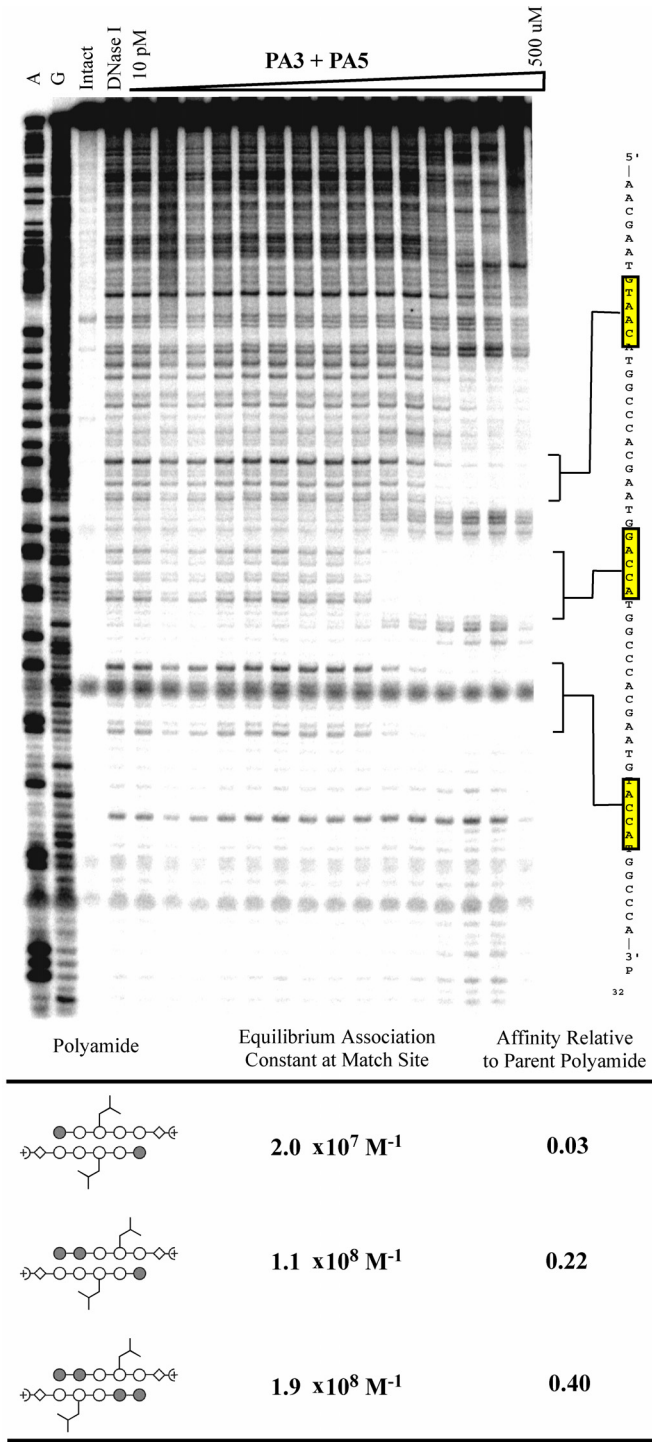
Figure 4.9. Top: Quantitative DNase footprinting experiment with ImPyPy(iBu)PyPy-β-Dp (**PA3**) and ImImPyPy(iBu)Py- β -Dp (**PA5**) on the 3'-^{32}P-labeled restriction fragment pDHN2. Gel lanes from left to right: A reaction, G reaction, intact DNA, DNase I standard, DNase digestion products in the presence of 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM polyamide. Bottom: Data table illustrating the binding affinities at each of the three match sites.
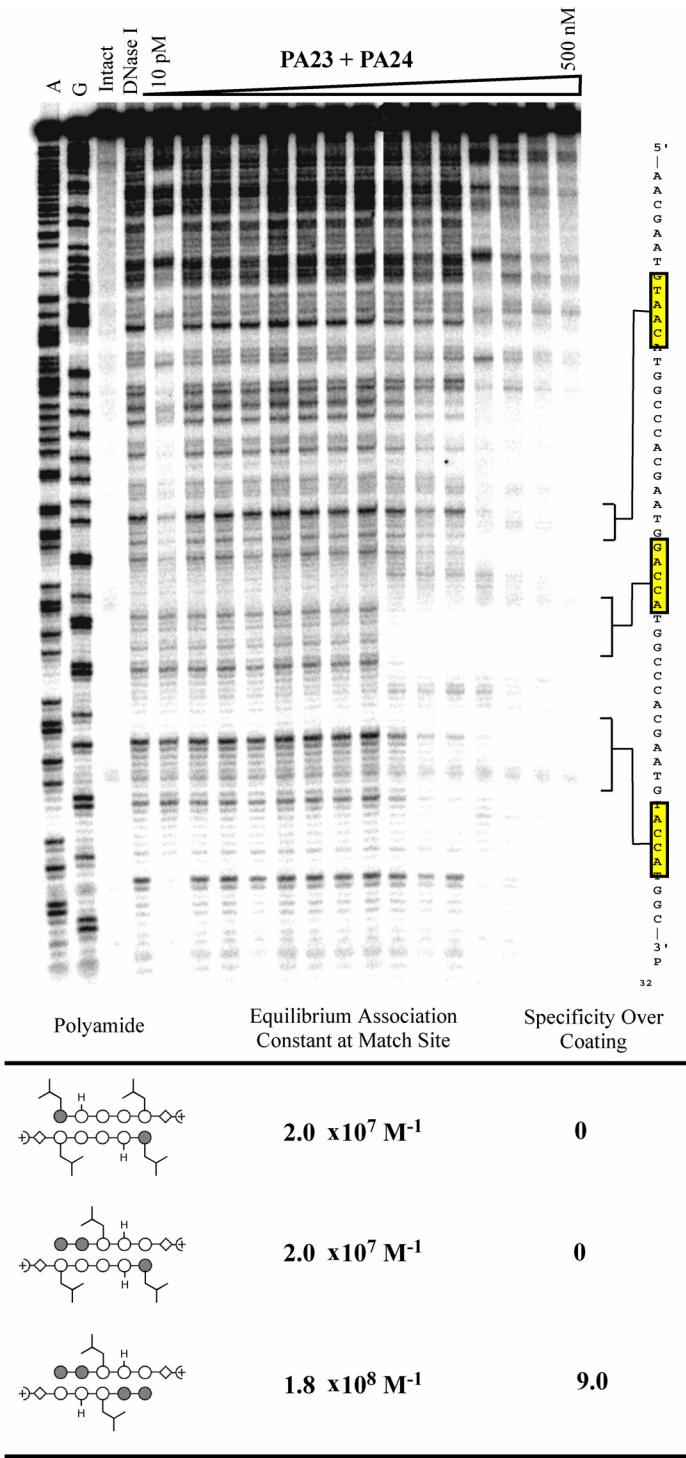
Figure 4.10. Top: Quantitative DNase footprinting experiment with Im(iBu)PyPyPyPy(iBu)-β-Dp (**PA23**) and ImImPy(iBu)PyPy- β -Dp (**PA24**) on the 3'-^{32}P-labeled restriction fragment pDHN2. Gel lanes from left to right: A reaction, G reaction, intact DNA, DNase I standard, DNase digestion products in the presence of 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM polyamide. Bottom: Data table illustrating the binding affinities at each of the three match sites.

in Figure 4.10, these polyamides bind each of the three designed sites with affinities below the parent polyamides. **PA23**, which contains only one isobutyl group, can still be seen to bind at its match site with an affinity of $1.8 \times 10^8 \text{ M}^{-1}$. No binding can be seen at either the one-imidazole or heterodimer match sites. Instead, protection from DNase I cleavage is incurred along the entire oligonucleotide (a phenomenon dubbed "coating") at concentrations as low as 50 nM. When **PA23** and **PA24** are footprinted separately, the coating is seen only when the DNA is equilibrated with **PA24**.

Discussion.

The results of the quantitative DNase I footprinting analysis have shown that the issue of hetero- versus homodimer formation can be controlled by non-covalent negative interactions. An example of this is the isobutyl side chain. When this side chain is paired across from itself in a DNA:polyamide complex, the polyamide affinity for the DNA is compromised 16-fold. By moving the sterically bulky groups apart by two residues, the binding affinity loss is rescued. Modeling data[15] for the isobutyl side chain homodimer pairing indicates that the bond between the pyrrole nitrogen and the isobutyl methylene group would need to bend 14° away from the plane of the pyrrole ring for the two bulky groups to fit within the minor groove together. Thus, the modeling data correlate well with the experimental footprinting data. When multiple isobutyl groups are placed on a single 5-ring polyamide (as in **PA24**), the polyamide loses its ability to discriminate DNA sequence and coats the oligonucleotide. One hypothesis to explain these results is that placement of multiple hydrophobic groups on a polyamide designed to bind in the hydrophobic minor groove causes favorable desolvation energetics to outcompete the

energy loss of binding the polyamide non-specifically to DNA. Thus, the polyamide is forced into the hydrophobic minor groove even if it is required to tolerate a mismatch at that site. Both the benzyl and ethyl benzyl side chains exhibit similar specific loss of affinity as the isobutyl side chain when homodimerically paired. Perhaps these side chains can be used as negative control elements without the unwanted coating behavior exhibited by polyamides with multiple isobutyl functionalities.

One interesting result is that the benzyl group (**PA20**) gave a significantly reduced binding affinity while the isosteric pentafluorobenzyl group (**PA17**) did not alter affinity. Modeling data[16] show that the benzyl group is unable to lie in the same plane as the polyamide without the benzyl 2H sterically clashing with the backbone carbonyl group of the polyamide chain. Two of these rotated benzyl groups cannot fit within the minor groove without disfavorable steric interactions when paired. The higher affinity of the pentafluorobenzyl group may be attributable to its quadrapole. Because the five electronegative fluorine atoms pull electron density out of the aromatic ring, the π cloud is significantly electropositive. The destabilizing steric effects of the ring/ring pairing may be balanced by a stabilizing electrostatic interaction between this electropositive moiety and the negatively charged phosphodiester backbone.

Unfortunately, non-covalent, positively interacting side chains fail to exhibit specific increases in dimer affinity. Each of the side chains carrying a positive charge shows an increase in affinity when incorporated into polyamides. However, the observed increase is not specific to a side chain/side chain pairing. The positive charge on the side chain is able to form a favorable electrostatic interaction with the negatively charged phosphodiester backbone of the DNA. The propylamino and guanidyl side chains give

larger binding affinity increases than the ethylamino-functionalized polyamide. This suggests that the longer chain length is more optimal for interacting with the phosphodiester backbone. Modeling shows that the ethylamino group lacks the chain length necessary to reach the DNA backbone, and therefore, would be an ideal candidate to participate in specific cation/π interactions while not participating in non-specific, associative DNA contacts. While two ethylamino groups present in a dimeric polyamide/DNA complex show only a modest 3-fold increase in affinity (due to non-specific electrostatic interactions with the DNA), pair-wise interactions with each of the aromatic side chains fail to exhibit affinity increases specific to the putative cation/π pairing.

Likewise, hydrogen bonding interactions between the primary amide groups and the N1 proton of pyrrole fail to impart specific affinity increases. Incorporation of the des-methyl pyrrole ring into the polyamide stabilizes the complex with DNA 3-5-fold. Thus, the apparent affinity increases seen when the des-methyl pyrrole is paired with the amide side chains are non-specific, and a result of the incorporation of the des-methyl ring, not a specific hydrogen bonding interaction.

Each of the non-covalent associative forces tested here have been shown to aid in protein folding and association. However, in proteins, these interacting side chains are removed from a protic solvent environment and interaction takes place inside the greasy protein interior, where their effects are increased. One hypothesis for the failure of non-covalent associative forces to incur specificity in this polyamide-based system is that the spatial area above the minor groove of DNA where the interactions tested here take place is a water-accessible area. This makes the associative interactions weaker due to water

solvation. An alternative hypothesis is that the polyamide/DNA complex is a flexible and dynamic complex, and that one or two additional weakly associative forces on the backbone are not enough to overcome the inherent motion of the polyamide strands with respect to the DNA. In conclusion, while non-covalent associative forces fail to specify for one binding mode, side chains such as the bulky isobutyl group are able to cause a specific destabilization of DNA binding, which may be used to refine the sequence space targeted by dimeric polyamide/DNA complexes.

Materials and Methods.

Dicyclohexylcarbodiimide (DCC), hydroxybezotriazole (HOBT), 2-(1H-benzotriazole-1-yl)-1,1,3,3,-tetramethyluronium hexafluorophosphate (HBTU), and 0.6 mmol/g Boc-β-Ala-Pam-Resin were purchased from Peptides International. 2-Bromoethyl acetamide was purchased from TCI America. All other chemicals were purchased from Aldrich Chemicals and used without further purification. Anhydrous dimethylformamide (DMF), N-methylpyrrolidone (NMP), N,N-dimethylpropylamine (Dp), and N,N-diisopropylethylamine (DIEA) were purchased from Aldrich Chemicals and stored over 3 Å molecular sieves. All other solvents were purchased from EM Sciences and were reagent-grade. Deuterated NMR solvents were purchased from Cambridge Isotopes.

^{1}H NMR spectra were recorded on a 300 MHz General Electric–QE NMR spectrometer with chemical shifts reported in parts per million relative to the residual solvent peak. UV spectra were recorded in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, LASER desorption/ionization time of

flight mass spectrometry was performed on a Voyager DE Pro Spectrometer. Electrospray ionization (E/I) mass spectrometry was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold System using a Rainen C_{18}, microsorb Mv, 5μm, 300 x 4.6 mm reversed-phase column in a 0.1%(wt/v) TFA aqueous solution with acetonitrile as eluent at a flow rate of 1.0 mL/min and a gradient elution of 1.25% acetonitrile/min. Preparative HPLC was performed on a Beckman Instrument using a Waters DeltaPak 25 x 100 mm, 100μm C_{18} reversed-phase column with a guard. The solvent was 0.1% (wt/v) aqueous TFA at 8.0 mL/min with 0.25%/min acetonitrile as the eluent. Gels were imaged using a Molecular Dynamics 400S PhosphorImager.

Restriction endonucleases, deoxyribonucleotide triphosphates, and glycogen were purchased from Boeringher-Mannheim. Sequenase was obtained from Amersham Life Sciences. DNase I and deproteinized calf thymus DNA were purchased from Pharmacia Biotech. [α-^{32}P]-Thymidine-5'-triphosphate (≥ 3000 Ci/mmol) and [α-^{32}P]-Deoxyadenosine-5'-triphosphate (≥ 6000 Ci/mmol) were purchased from New England Nucleosides. Water was used from a Millipore Milli-Q purification system. All buffer reagents were purchased from Fluka Biochemika Microselect. All buffers were sterilized by filtration through Nalgene 0.2 μm cellulose nitrate filtration devices.

Monomer Synthesis.

1,2,3-Benzotriazol-1-yl 4-[(tert-butoxy)carbonylamino]-1-methylpyrrole-2-carboxylic acid (Boc-Py-OBt), 4[(tert-butoxy)carbonylamino]-1-methylimidazole-2-carboxylic acid (Boc-Im-OH), imidazole-2-carboxylic acid (Im-OH), 4[(tert-

butoxy)carbonylamino]-pyrrole-2-carboxylic acid (Boc-Ds-OH, **7j**), and 4-[(tert-butoxy)carbonylamino]-1-[3-(1,3-dioxoisoindolin-2-yl)propyl]pyrrole-2-carboxylic acid (Boc-Py(3Ph)-OH, **7a)** were synthesized according to published procedures.[11-13] 4-[(tert-butoxy)carbonylamino)-1benzylpyrrole-2-carboxylic acid (Boc-Py(Bz)-OH, **7i**) was prepared by Nick Wurtz (unpublished results).

General procedure for alkylation of 3,3-dimethyl-3-silabutyl 4-nitropyrrole-2-carboxylate (2) and ethyl 4-nitropyrrole-2-carboxylate (1) for compounds 5a, d–f, h, k

To a solution of the 4-nitropyrrole-2-carboxylate (1.0 mmol) in 25 mL acetone dried over K_2CO_3 was added powdered K_2CO_3 (2.0 mmol). This suspension was stirred at rt for 30 min.   The appropriate alkyl bromide (1.2 mmol) was then added, followed by tetrabutylammonium iodide (0.2 mmol). The reaction was refluxed until completion as determined by TLC (1–24h). The reaction was cooled and the acetone evaporated *in vacuo*. The residue was taken up in water (50 mL) and extracted with ethyl acetate (2 x 100 mL). The combined organic layer was dried (sodium sulfate) and concentrated *in vacuo*. The residue was then chromatographed in the appropriate solvent conditions to yield the pure 1-alkylated 4-nitropyrrole-2-carboxylate.

General procedure for alkylation of 3,3-dimethyl-3-silabutyl 4-[(tert-butyl)carbonylamino]pyrrole-2-carboxylate (4) and ethyl 4-[(tert-butyl)carbonylamino]pyrrole-2-carboxylate (3) for compounds 6b, c

To a solution of the 4-[(tert-butoxy)carbonylamino]pyrrole-2-carboxylate (1.0 mmol) in 25 mL acetone dried over K_2CO_3 was added powdered K_2CO_3 (2.0 mmol). This suspension was stirred at rt for 30 min. The appropriate alkyl bromide (1.2 mmol) was then added, followed by tetrabutylammonium iodide (0.2 mmol). The reaction was refluxed until completion as determined by TLC (1-24h). The reaction was cooled and the acetone evaporated *in vacuo*. The residue was taken up in water (50 mL) and extracted with ethyl acetate (2 x 100 mL). The combined organic layer was dried (sodium sulfate) and concentrated *in vacuo*. The residue was then chromatographed in the appropriate solvent conditions to yield the pure 1-alkylated 4-[(tert-butoxy)carbonylamino]-2-carboxylate.

General procedure for reduction and boc-protection of alkylated3,3-dimethyl-3-silabutyl 4-nitropyrrole-2-carboxylate and ethyl 4-nitropyrrole-2-carboxylate derivatives (5a,d-f,h,k) for compounds 6a, d–f, h, k

To a solution of the 4-nitropyrrole 2-carboxylate derivative (1.0 mmol) in 25 mL ethyl acetate was added 10% Pd/C (20% by wt). The mixture was stirred vigorously under 400 psi of hydrogen for 1–12 h. When complete, the reaction was filtered through Celite, and the Celite washed with ethyl acetate (3 x 10 mL). The filtrates were combined and Boc anhydride (1.1 mmol) was added, followed by 1M $NaHCO_3$ (50 mL). The biphasic reaction mixture was stirred vigorously at rt until completion as determined by TLC. The layers were separated and the organic layer washed with 10% citric acid (2 x 50 mL) followed by brine (2 x 50 mL). The organic layer was dried (sodium sulfate) and concentrated *in vacuo*. The residue was then column chromatographed in the appropriate

solvent system to yield the pure 4-[(tert-butoxy)carbonylamino]pyrrole-2-carboxylate derivative.

General procedure for the hydrolysis of ethyl esters 6c, d, f, h, k

To a solution of the 4-[(tert-butoxy)carbonylamino]pyrrole-2-carboxylate derivative (1.0 mmol) in 10 mL ethanol was added 10 mL of a 5.0 M solution of sodium hydroxide. The reaction was stirred at rt until completion as determined by TLC. The ethanol was removed *in vacuo* and the aqueous layer acidified with 10% H_2SO_4 to pH 3. The carboxylic acid product was then extracted with ethyl acetate (2 x 25 mL). The combined organic layer was dried (sodium sulfate) and concentrated *in vacuo* to yield the pure product.

General procedure for the deprotection of trimethylsilylethyl esters 6a, b, e

A solution of the 3,3-dimethyl-3-silabutyl 4-[(tert-butoxy)carbonylamino]pyrrole-2-carboxylate derivative (1.0 mmol) in 5 mL anhydrous THF under argon was cooled to 0 °C on an ice bath. A 2.0 M solution of TBAF in THF (0.5 mL) was added drop-wise to the cooled reaction mixture. The reaction was allowed to warm slowly to room temp where it was stirred overnight under argon. The reaction was then quenched with the addition of 1 mL 10% citric acid and cooled to -20 °C where the pure carboxylic acid crystallized out. The crystalline product was filtered and washed with cold hexanes to yield the pure product.

3,3-dimethyl-3-silabutyl-4-[(tert-butoxy)carbonylamino]-1-[(2,3,4,5,6-pentafluorophenyl)methyl]pyrrole-2-carboxylate (6b)

From 2.0 g (7.8 mmol) **4** and 2,3,4,5,6-pentafluorobenzyl bromide (2.04 g, 7.8 mmol) was yielded 3.2 g (88%) of **6b** as a pale brown solid after flash chromatography in 1:1 hexanes/ethyl acetate. TLC (ethyl acetate) R_f 0.85 ^{1}H NMR(DMSO-d_6) δ 9.20 (s, 1H), 7.17 (s, 1H), 6.60 (s, 1H), 5.68 (s, 2H), 4.23 (t, 2H, J=8.4), 1.42 (s, 9H), 0.95 (t, 2H, J=8.4), 0.00 (s, 9H). E/I MS m/e 507.4 (M+H) (506.17 calcd. for $C_{22}H_{27}F_5N_2O_4Si$).

4-[(tert-butoxy)carbonylamino]-1-[(2,3,4,5,6-pentafluorophenyl)methyl]pyrrole-2-carboxylic acid (Boc-Py(5F)-OH) (7b)

From 3.0 g (5.94 mmol) **6b** was yielded 2.1 g (88%) of **7b** as a pale brown powder. TLC (ethyl acetate) R_f 0.05 ^{1}H NMR(DMSO-d_6) δ 12.22 (bs, 1H) 9.16 (s, 1H), 7.09 (s, 1H), 6.59 (s, 1H), 5.65 (s, 2H), 1.41 (s, 9H). E/I MS m/e 405.2(M-H)(406.10 calcd. for $C_{17}H_{15}F_5N_2O_4$).

Ethyl 4-[(tert-butoxy)carbonylamino]-1-(2-phenylethyl)pyrrole-2-carboxylate (6c)

From 6.0g (23.7 mmol) **3** and bromoethyl benzene (5.24g, 28.3 mmol) was yielded 3.6 g (42%) **6c** after flash chromatography in 3:17 ethyl acetate/hexanes as a clear oil. TLC (1:9 hexanes/ethyl acetate) R_f 0.17 ^{1}H NMR(DMSO-d_6) δ 9.11 (s, 1H), 7.30 (m, 1H), 7.26 (d, 2H, J=6.9), 7.19 (d, 2H, J=6.9), 7.11 (s, 1H), 6.64 (s, 1H), 4.39 (t, 2H, J=8.1), 4.17 (q, 2H, J=7.2), 2.90 (t, 2H, J=7.4), 1.42 (s, 9H), 1.24 (t, 2H, J=7.2). E/I MS m/e 359.3 (M+H) (359.19 calcd. for $C_{20}H_{26}N_2O_4$).

4-[(tert-butoxy)carbonylamino]-1-(2-phenylethyl)pyrrole-2-carboxylic acid (Boc-Py(2Bz)-OH) (7c)

From 1.5 g (4.2 mmol) **6c** was yielded 1.1 g (79%) of **7c** as a white powder. TLC (ethyl

acetate) R_f 0.05 ^{1}H NMR(DMSO-d_6) δ 8.89 (s, 1H), 7.26 (m, 5H), 6.85 (s, 1H), 6.36 (s,

1H), 4.42 (t, 2H, J=7.8), 2.90 (t, 2H, J=7.6), 1.41 (s, 9H). E/I MS m/e 329.1 (M-H)

(330.16 calcd. for $C_{18}H_{22}N_2O_4$).

Ethyl 4-nitro-1-(2-carbamoylethyl)pyrrole-2-carboxylate (5d)

From 2.5 g (13.6 mmol) **1** and 2-bromoethyl acetamide (2.27 g, 15 mmol) was yielded

2.8 g (81%) of **5d** after flash chromatography in 2:1 ethyl acetate/hexanes as a white

powder. TLC (2:1 hexanes/ethyl acetate) R_f 0.22 ^{1}H NMR(DMSO-d_6) δ 8.13 (d, 1H,

J=1.8), 7.35 (s, 1H), 7.32 (d, 1H, J=1.8), 6.92 (s, 1H), 4.52 (t, 2H, J=6.9), 4.32 (q, 2H,

J=7.2), 2.59 (t, 2H, J=6.6), 1.29 (t, 2H, J=7.5). E/I MS m/e 256.4 (M+H) (255.09 calcd.

for $C_{10}H_{13}N_3O_5$).

Ethyl 4-[(tert-butoxy)carbonylamino]-1-(2-carbamoylethyl)pyrrole-2-carboxylate

(6d)

From 2.5 g (11 mmol) **5d** was yielded 2.14 g (60%) **6d** after flash chromatography in

100% ethyl acetate as a white powder. TLC (ethyl acetate) R_f 0.71 ^{1}H NMR(DMSO-d_6)

δ 9.12 (s, 1H), 7.32 (s, 1H), 7.04 (s, 1H), 6.81 (s, 1H), 6.60 (s, 1H), 4.39 (t, 2H, J=6.6),

4.18 (q, 2H, J=7.4), 2.42 (t, 2H, J=6.6), 1.41 (s, 9H), 1.27 (t, 2H, J=7.4). E/I MS m/e

326.2 (M+H) (325.16 calcd. for $C_{15}H_{23}N_3O_5$).

4-[(tert-butoxy)carbonylamino]-1-(2-carbamoylethyl)pyrrole-2-carboxylic

acid (Boc-Py(2Am)-OH) (7d)

From 600 mg (1.85 mmol) **6d** was yielded 550 mg (97%) **7d** as a white powder. TLC (ethyl acetate) R_f 0.05 ^{1}H NMR(DMSO-d_6) δ 9.05 (s, 1H), 7.30 (s, 1H), 7.03 (s, 1H), 6.82 (s, 1H), 6.57 (s, 1H), 4.36 (t, 2H, J=6.9), 2.45 (t, 2H, J=6.8), 1.42 (s, 9H). E/I MS m/e 296.1 (M-H) (297.13 calcd. for $C_{13}H_{19}N_3O_5$).

3,3-dimethyl-3-silabutyl-4-nitro-1-[2-(1,3-dioxoisoindolin-2-yl)ethyl]pyrrole-2-carboxylate (5e)

2 (2.0 g, 7.8 mmol) and 2-bromoethyl phthalimide (2.39 g, 9.37 mmol) were combined in acetone along with powdered K_2CO_3 (1.61 g, 11.7 mmol) and tetrabutylammonium iodide (576 mg, 1.5 mmol) and stirred at 40 °C for 48 h. Every 12 h, an additional 0.5 equivalents of 2-bromoethyl phthalimide (994 mg, 3.4 mmol) was added. 3.0 g (90%) **5e** was isolated after the standard workup and flash chromatography in 4:1 hexanes/ethyl acetate as a clear oil. TLC (6:1 hexanes/ethyl acetate) R_f 0.12 ^{1}H NMR(DMSO-d_6) δ 8.38 (d, 1H, J=2.4), 7.84 (m, 4H), 7.22 (d, 1H, J=2.4), 4.64 (t, 2H, J=6.6), 4.04 (t, 2H, J=7.4), 3.96 (t, 2H, J=6.6), 0.79 (t, 2H, J=7.4), 0.00 (s, 9H). E/I MS m/e 430.5 (M+H) (429.14 calcd. for $C_{20}H_{23}N_3O_6Si$).

3,3-dimethyl-3-silabutyl 4-[(tert-butoxy)carbonylamino]-1-[2-(1,3-dioxoisoindolin-2-yl)ethyl]pyrrole-2-carboxylate (6e)

From 1.25 g (2.9 mmol) **5e** was isolated 1.3 g (90%) **6e** after flash chromatography in 5:2 hexanes/ethyl acetate. TLC (1:1 hexanes/ethyl acetate) R_f 0.86 ^{1}H NMR(DMSO-d_6) δ 9.10 (s, 1H), 7.83 (m, 4H), 7.04 (s, 1H), 6.56 (s, 1H), 4.51 (t, 2H, J=6.6), 3.93 (t, 2H,

J=6.6), 3.87 (t, 2H, J=7.4), 1.42(s, 9H), 0.76 (t, 2H, J=7.4), 0.00 (s, 9H). E/I MS m/e 500.1 (M+H) (499.21 calcd. for $C_{25}H_{33}N_3O_6Si$).

4-[(tert-butoxy)carbonylamino]-1-[2-(1,3-dioxoisoindolin-2-yl)ethyl]pyrrole-2-carboxylic acid (Boc-Py(2P)-OH) (7e)

From 1.15 g (2.3 mmol) **6e** was yielded 900 mg (98%) **7e** as a white powder. TLC (2:1 hexanes/ethyl acetate) R_f 0.05 ^{1}H NMR(DMSO-d_6) δ 12.05 (bs, 1H), 8.98 (s, 1H), 7.79 (m, 4H), 6.88 (s, 1H), 6.51 (s, 1H), 4.44 (t, 2H, J=6.6), 3.87 (t, 2H, J=6.6), 1.36 (s, 9H). E/I MS m/e 398.6 (M-H) (399.14 calcd. for $C_{20}H_{21}N_3O_6$).

Ethyl 4-nitro-1-(2-methylpropyl)pyrrole-2-carboxylate (5f)

From 6.0 g (32.6 mmol) **1** and 1-bromo-2-methylpropane (5.8 g, 42.4 mmol) was yielded 7.14 g (91%) **5f** as a yellow solid. TLC (3:2 hexanes/ethyl acetate) R_f 0.90 ^{1}H NMR(DMSO-d_6) δ. E/I MS m/e 241.1 (M+H) (240.11 calcd. for $C_{11}H_{16}N_2O_4$).

Ethyl 4-[(tert-butoxy)carbonylamino]-1-(2-methylpropyl)pyrrole-2-carboxylate (6f)

From 3.5 g (14.5 mmol) **5f** was yielded 4.1 g (91%) **6f** after flash chromatography in 5:2 hexanes/ethyl acetate as an off-white solid. TLC (3:2 hexanes/ethyl acetate) R_f 0.80 ^{1}H NMR(DMSO-d_6) δ 9.17 (s, 1H), 7.06 (s, 1H), 6.64 (s, 1H), 4.16 (q, 2H, J=6.0), 4.01 (d, 2H, J=7.2), 1.91 (m, 1H), 1.42 (s, 9H), 1.23 (t, 2H, J=5.9), 0.77 (d, 6H, J=7.2). E/I MS m/e 310.9 (M+H) (310.19 calcd. for $C_{16}H_{26}N_2O_4$).

4-[(tert-butoxy)carbonylamino]-1-(2-methylpropyl)pyrrole-2-carboxylic acid (Boc-Py(iBu)-OH) (7f)

From 1.24 g (3.99 mmol) **6f** was yielded 980 mg (88%) **7f** as a white solid. TLC (3:2 hexanes/ethyl acetate) R_f 0.05 ^{1}H NMR(DMSO-d_6) δ 9.03 (s, 1H), 7.00 (s, 1H), 6.56 (s, 1H), 4.03 (d, 2H, J=6.0), 1.91 (m, 1H), 1.41 (s, 9H), 0.78 (d, 6H, J=7.2). E/I MS m/e 281.3 (M-H) (282.16 calcd. for $C_{14}H_{22}N_2O_4$).

Tert-butyl 3-(2-[4-nitro-2-(ethoxycarbonyl)pyrrolyl]ethyl)indolecarboxylate (5k)

From 300 mg (1.63 mmol) **1** and tert-butyl 3-(bromomethyl)indolecarboxylate (554 mg, 1.79 mmol) was yielded 485 mg (73%) **5k** after flash chromatography in 6:1 hexanes/ethyl acetate as a pale yellow powder. TLC (2:1 hexanes/ethyl acetate) R_f 0.95 ^{1}H NMR(DMSO-d_6) δ 8.405 (d, 1H, J=2.1), 8.03 (d, 1H, J=8.4), 7.69 (s, 1H), 7.63 (d, 1H, J=8.4), 7.34 (d, 1H, J=2.1), 7.25 (m, 2H), 5.74 (s, 2H), 4.27 (q, 2H, J=7.2), 1.60(s, 9H), 1.27 (t, 2H, J=7.2). E/I MS m/e 414.2 (M+H) (413.16 calcd. for $C_{21}H_{23}N_3O_6$).

Tert-butyl-3-(2-{4-[(tert-butoxy)carbonylamino]-2-(ethoxycarbonyl)pyrrolyl)ethyl)indolecarboxylate (6k)

From 450 mg (1.09 mmol) **5k** was yielded 450 mg (86%) **6k** after flash chromatography in 6:1 hexanes/ethyl acetate as a pale yellow solid. TLC (5:1 hexanes/ethyl acetate) R_f 0.37 ^{1}H NMR(DMSO-d_6) δ 9.11 (s, 1H), 8.01 (d, 1H, J=2.4), 7.55 (s, 1H), 7.54 (d, 1H, J=9.0), 7.32 (t, 1H, J=8.1), 7.21 (s, 1H), 7.20 (m, 1H), 6.65 (s, 1H), 5.62 (s, 2H), 4.20 (q, 2H, J=7.2), 1.60(s, 9H), 1.38 (s, 9H), 1.24 (t, 2H, J=6.9). E/I MS m/e 484.3 (M+H) (483.24 calcd. for $C_{26}H_{33}N_3O_6$).

4-[(tert-butoxy)carbonylamino]-1-(2-indole-2-ylmethyl)pyrrole-2-carboxylic

acid (Boc-Py(In)-OH) (7g)

From 400 mg (0.83 mmol) **6k** was yielded 286 mg (97%) **7g** as a pale yellow solid. TLC (7:1 hexanes/ethyl acetate) R_f 0.10 ^{1}H NMR(DMSO-d_6) δ 11.03 (s, 1H), 9.11 (s, 1H), 7.49 (t, 1H, J=6.6), 7.35 (d, 1H, J=8.1), 7.33 (s, 1H), 7.06 (s, 1H), 7.03 (m, 1H),6.95 (t, 1H, J=7.2), 6.59 (s, 1H), 5.61 (s, 2H), 1.37 (s, 9H). E/I MS m/e 354.5 (M-H) (355.15 calcd. for $C_{19}H_{21}N_3O_4$).

Ethyl 4-nitro-1-(cyanomethyl)pyrrole-2-carboxylate (5h)

Sodium Hydride (60% dispersion in mineral oil, 304 mg, 7.6 mmol) was dissolved in 3 mL anhydrous DMF and stirred under argon at 0 °C. **1** (1.0 g, 5.43 mmol) dissolved in 6 mL anhydrous DMF was then added drop wise and the solution stirred for 30 min. Bromoacetonitrile (1.15 g, 9.77 mmol) was then added drop wise to the reaction mixture allowed to warm to room temperature, where it was stirred overnight. The reaction mixture was poured into water (50 ml) and the product extracted with ethyl acetate (2 x 100 mL). The combined organic layer was then washed with brine (2 x 50 mL), dried over sodium sulfate, and concentrated *in vacuo*. The residue was then flash chromatographed in 1:1 ethyl acetate/hexanes to yield 1.07 g (89%) **5h** as a pale yellow oil. TLC (1:1 hexanes/ethyl acetate) R_f 0.60 ^{1}H NMR(DMSO-d_6) δ 8.40 (d, 1H, J=2.4), 7.41 (d, 1H, J=2.4), 5.50 (s, 2H), 4.31 (q, 2H, J=6.9), 1.30 (t, 2H, J=6.9). E/I MS m/e 224.0 (M+H) (223.06 calcd. for $C_9H_9N_3O_4$).

Ethyl 4-[(tert-butoxy)carbonylamino]-1-(cyanomethyl)pyrrole-2-carboxylate (6h)

From 1.07 g (4.9 mmol) **5h** was yielded 560 mg (40%) **6h** after flash chromatography in 7:2 hexanes/ethyl acetate as a white powder. TLC (8:5 hexanes/ethyl acetate) R_f 0.56 ^{1}H NMR(DMSO-d_6) δ 9.26 (s, 1H), 7.31 (s, 1H), 6.70 (s, 1H), 5.40 (s, 2H), 4.20 (q, 2H, J=6.9), 1.43 (s, 9H), 1.26 (t, 2H, J=6.9). E/I MS m/e 294.4 (M+H) (293.14 calcd. for $C_{14}H_{19}N_3O_4$).

4-[(tert-butoxy)carbonylamino]-1-(cyanomethyl)pyrrole-2-carboxylic acid (Boc-Py(CN)-OH) (7h)

From 340 mg (1.16 mmol) **6h** was yielded 300 mg (97%) **7h** as a white solid. TLC (1:1 hexanes/ethyl acetate) R_f 0.05 ^{1}H NMR(DMSO-d_6) δ 12.19 (bs, 1H), 9.09 (s, 1H), 7.09 (s, 1H), 6.57 (s, 1H), 4.92 (s, 2H), 1.42 (s, 9H). E/I MS m/e 264.0 (M-H) (265.11 calcd. for $C_{12}H_{15}N_3O_4$).

2,2,2-trichloro-1-[1-(2-methylpropyl)imidazol-2-yl]ethan-1-one (Im(iBu)-CK) (9)

(2-methylpropyl)imidazole (**8**) (10.0 g, 80.6 mmol) was mixed with 32 mL dichloromethane and added drop wise to a stirring solution of trichloroacetylchloride (14.74 g, 81.5 mmol) in 48 mL dichloromethane. The resulting solution was stirred at rt for 2 h. The reaction mixture was cooled to 0 °C and triethylamine (8.15 g, 80.6 mmol) was added drop wise. The triethylammonium chloride salt was filtered off and the mother liquor was concentrated *in vacuo*. The crude product was purified by flash chromatography in 2:1 hexanes/ethyl acetate to yield **9** (17.4 g, 84%) as a white powder. TLC (2:1 hexanes/ethyl acetate) R_f 0.46 ^{1}H NMR(DMSO-d_6) δ 7.77 (s, 1H), 7.36 (s, 1H),

4.23 (d, 2H, J=7.4), 1.95 (m, 1H), 0.83 (d, 6H, J=6.6). E/I MS m/e 269.1 (M+H) (267.99 calcd. for $C_9H_{11}Cl_3N_2O$).

2,2,2-trichloro 4-nitro-1-[1-(2-methylpropyl)imidazol-2-yl]ethan-1-one (10)

Compound **9** (1.0 g, 3.6 mmol) was dissolved in 10 mL acetic anhydride and cooled to an internal temperature of 0 °C. Fuming nitric acid (1.5 mL) was then added drop wise, followed by 0.5 mL concentrated H_2SO_4. The reaction mixture was slowly allowed to warm to room temperature where it was stirred overnight. After 10 h, the reaction was taken up in 25 mL water and the product extracted with ethyl acetate (2 x 50 mL). The combined organic layers were dried over sodium sulfate and concentrated *in vacuo*. The residue was then flash chromatographed in 4:1 hexanes/ethyl acetate to yield **10** (756 mg, 64%) as a pale yellow solid. TLC (3:1 hexanes/ethyl acetate) R_f 0.69 ^{1}H NMR(DMSO-d_6) δ 8.61 (s, 1H), 4.25 (d, 2H, J=7.2), 2.07 (m, 1H), 0.82 (d, 6H, J=6.6). E/I MS m/e 314.0 (M+H) (312.98 calcd. for $C_9H_{10}Cl_3N_3O_3$).

Ethyl 1-(2-methylpropyl)-4-nitroimidazole-2-carboxylate (11)

Compound **10** (450 mg, 1.38 mmol) was slurried with 5 mL ethanol and sodium hydride (60% dispersion in mineral oil, 100 mg) dissolved in 3 mL ethanol was added drop-wise over 10 min. The suspension was then heated to reflux for 30 min. The reaction was allowed to cool to rt and then was neutralized with cold 1 N HCl to pH7. The product was then extracted with chloroform (2 x 15 mL). Compound **11** (200 mg, 68%) was isolated without further purification as a clear oil. TLC (3:1 hexanes/ethyl acetate) R_f 0.31 ^{1}H NMR(DMSO-d_6) δ 8.56 (s, 1H), 4.25 (d, 2H, J=7.2), 3.41 (q, 2H, J=6.6), 2.08

(m, 1H), 1.03 (t, 2H, J=6.6), 0.83 (d, 6H, J=6.0). E/I MS m/e 242.1 (M+H) (241.11 calcd. for $C_{10}H_{15}N_3O_4$).

Polyamide Synthesis.

The parent polyamides ImPyPyPyPy-β-Dp (**PA1**) and ImImPyPyPy-β-Dp (**PA2**) were prepared according to standard solid-phase protocols.[13]

ImPyPy(Bz)PyPy-β-Dp (**PA18**) The manual solid-phase synthesis was performed as in previously published work.[13] Boc-β-Ala-Pam-Resin (250 mg, 0.25 mmol) was swelled in DMF inside a 20 mL glass reaction vessel fitted with a glass filter and stopcock. The reaction vessel was drained after five minutes, the resin washed with 2 x 10 mL DCM, and deprotected with 50% TFA:DCM (10 mL). The vessel was shaken for 30 minutes and then drained and the resin washed with 3 x 10 mL DCM, 1 x 10 mL 4:1 DMF:DIEA, 1 x 10 mL DMF. Boc-Py-OBt (178 mg, 0.5 mmol) was added to the resin along with 1 mL NMP and 0.25 mL DIEA, and the reaction shaken at room temperature for two hours. A resin sample (4 mg) was taken out and cleaved in Dp at 90 °C for 10 minutes and used for HPLC analysis to verify completion of the coupling. 1 mL acetic anhydride was then added to the reaction and shaken for 10 minutes. The cycle was repeated for each new monomer using two equivalents of each Boc-protected monomer. Boc-Py(Bz)-OH, Boc-Ds-OH, Im-OH, and Boc-Im-OH were first activated with HBTU (1.0 equivalent) in NMP (1 mL) and DIEA (0.25 mL) for 5 minutes at room temperature. Im(iBu)-CCl$_3$ was added directly to the resin with 1 mL NMP and 0.25 mL DIEA and shaken at 37 °C for 2 h. When the polyamide synthesis was complete, the vessel was drained and washed with 2 x 10 mL each DMF, DCM, MeOH, and diethylether. Polyamide was cleaved from

resin in a glass vial with 2 mL Dp at 55 °C for 12 h. The cleavage product was purified by reversed phase HPLC. UV λmax (H20) (ε) 312 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 10.46 (s, 1H), 10.05 (s, 2H), 9.87 (s, 1H), 9.26 (bs, 1H), 8.02 (m, 2H), 7.38 (s, 1H), 7.34 (s, 1H), 7.27 (m, 3H), 7.18 (s, 1H), 7.16 (s, 1H), 7.13 (s, 2H), 7.10 (s, 2H), 7.03 (s, 1H) 6.85 (s, 1H), 5.59 (s, 2H), 3.97 (s, 3H), 3.82 (s, 3H), 3.81 (s, 3H), 3.77 (s, 3H), 3.35 (q, 2H, J=6.9), 3.08 (q, 2H, J=6.6), 2.98 (m, 2H), 2.72 (s, 3H), 2.71 (s, 3H), 2.32 (t, 2H, J=7.2), 1.71 (m, 2H). MALDI-TOF MS 846.34 (M+H) (845.42 calcd. for C$_{43}$H$_{51}$N$_{13}$O$_6$).

ImPyPyPyPy-β-Dp (PA1) UV λmax (H20) (ε) 310 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 10.47 (s, 1H), 9.56 (s, 1H), 9.93 (s, 1H), 9.89 (s, 1H), 9.20 (bs, 1H), 8.03 (m, 2H), 7.40 (s, 1H), 7.28 (s, 1H), 7.22 (m, 2H), 7.16 (s, 2H), 7.16 (s, 1H), 7.13 (s, 1H), 7.05 (s, 2H), 6.87 (s, 1H), 3.99 (s, 3H), 3.85 (s, 6H), 3.82 (s, 3H), 3.79 (s, 3H), 3.39 (q, 2H, J=6.0), 3.10 (q, 2H, J=6.0), 2.99 (m, 2H), 2.74 (s, 3H), 2.72 (s, 3H), 2.34 (t, 2H, J=7.2), 1.72 (m, 2H). MALDI-TOF MS 770.7 (M+H) (769.37 calcd. for C$_{36}$H$_{45}$N$_{13}$O$_6$).

ImImPyPyPy-β-Dp (PA2) UV λmax (H20) (ε) 310 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 10.35 (s, 1H), 9.96 (s, 1H), 9.89 (s, 1H), 9.76 (s, 1H), 9.20 (bs, 1H), 8.03 (m, 2H), 7.57 (s, 1H), 7.46 (s, 1H), 7.27 (d, 1H, J=1.8), 7.22 (d, 1H, J=1.7), 7.15 (s, 2H), 7.07 (s, 1H), 7.05 (d, 1H, J=1.8), 6.86 (d, 1H, J=1.7), 4.00 (s, 6H), 3.84 (s, 3H), 3.83 (s, 3H), 3.79 (s, 3H), 3.38 (q, 2H, J=6.0), 3.10 (q, 2H, J=5.7), 2.90 (m, 2H), 2.74 (s, 3H), 2.72 (s, 3H), 2.33 (t, 2H, J=7.8), 1.72 (m, 2H). MALDI-TOF MS 771.4 (M+H) (770.37 calcd. for C$_{36}$H$_{45}$N$_{13}$O$_6$).

ImPyPy(iBu)PyPy-β-Dp (**PA3**) UV λmax (H20) (ε) 311 nm (43,450). ^{1}H NMR (DMSO-d6) δ 10.47 (s, 1H), 9.96 (s, 1H), 9.93 (s, 1H), 9.89 (s, 1H), 9.25 (bs, 1H), 8.03 (m, 2H), 7.39 (s, 1H), 7.28 (s, 1H), 7.24 (s, 1H), 7.19 (s, 1H), 7.15 (m, 3H), 7.04 (s, 2H), 6.86 (s, 1H), 4.11 (d, 2H, J=6.0), 3.98 (s, 3H), 3.84 (s, 3H), 3.82 (s, 3H), 3.78 (s, 3H), 3.36 (q, 2H, J=6.6), 3.09 (q, 2H, J=6.0), 2.98 (m, 2H), 2.73 (s, 3H), 2.71 (s, 3H), 2.33 (t, 2H, J=6.6), 1.97 (m, 1H), 1.71 (m, 2H), 0.79 (d, 6H, J=6.6). MALDI-TOF MS 812.5 (M+H) (811.43 calcd. for $C_{40}H_{53}N_{13}O_6$).

ImImPy(iBu)PyPy-β-Dp (**PA4**) UV λmax (H20) (ε) 311 nm (43,450). ^{1}H NMR (DMSO-d6) δ 10.34 (s, 1H), 9.96 (s, 1H), 9.89 (s, 1H), 9.74 (s, 1H), 9.31 (bs, 1H), 8.04 (m, 2H), 7.56 (s, 1H), 7.45 (s, 1H), 7.30 (s, 1H), 7.20 (s, 1H), 7.15 (s, 1H), 7.10 (s, 1H), 7.07 (s, 1H), 7.03 (s, 1H), 6.86 (s, 1H), 4.11 (d, 2H, J=6.6), 3.99 (s, 6H), 3.82 (s, 3H), 3.80 (s, 3H), 3.36 (q, 2H, J=6.6), 3.09 (q, 2H, J=6.6), 2.97 (m, 2H), 2.72 (s, 3H), 2.71 (s, 3H), 2.32 (t, 2H, J=6.6), 1.96 (m, 1H), 1.72 (m, 2H), 0.78 (d, 6H, J=5.6) MALDI-TOF MS 813.45 (M+H) (812.43 calcd. for $C_{39}H_{52}N_{14}O_6$).

ImImPyPy(iBu)Py-β-Dp (**PA5**) UV λmax (H20) (ε) 311 nm (43,450). ^{1}H NMR (DMSO-d6) δ 10.34 (s, 1H), 9.96 (s, 1H), 9.88 (s, 1H), 9.75 (s, 1H), 9.30 (bs, 1H), 8.04 (m, 2H), 7.56 (s, 1H), 7.45 (s, 1H), 7.27 (s, 1H), 7.23 (s, 1H), 7.15 (s, 1H), 7.14 (s, 1H), 7.07 (s, 1H), 7.02 (s, 1H), 6.84 (s, 1H), 4.10 (d, 2H, J=7.2), 3.99 (s, 6H), 3.84 (s, 3H), 3.78 (s, 3H), 3.35 (q, 2H, J=6.0), 3.09 (q, 2H, J=6.3), 2.98 (m, 2H), 2.72 (s, 3H), 2.71 (s, 3H), 2.33 (t, 2H, J=7.2), 1.96 (m, 1H), 1.71 (m, 2H), 0.78 (d, 6H, J=6.3). MALDI-TOF MS 813.45 (M+H) (812.43 calcd. for $C_{39}H_{52}N_{14}O_6$).

ImPy(Am)PyPyPy-β-Dp (**PA6**) This polyamide was made by two distinct methods. In the first method, Each monomeric unit was coupled using the standard solid phase methodology. The Py(Am) unit was incorporated as the Boc-Py(CN)-OH (**7h**) which was hydrated by trace amounts of water in the subsequent acidic deprotection steps. Alternatively, **PA10** (10 μMol) was dissolved in 0.5 mL anhydrous DMF. NaH (80 μMol) was then added and the reaction allowed to sit at 0 °C for 15 minutes. Bromoacetamide (10 μMol) was then added and the reaction monitored by rp-HPLC. After 10 minutes, the reaction was complete and was quenched by the addition of 7.5 mL 0.1% TFA and purified by reversed-phase HPLC (50% yield). UV λmax (H20) (ε) 311 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 10.49 (s, 1H), 9.97 (s, 1H), 9.94 (s, 1H), 9.88 (s, 1H), 9.28 (bs, 1H), 8.03 (m, 2H), 7.40 (s, 1H), 7.29 (s, 1H), 7.28 (s, 1H), 7.21 (m, 3H), 7.15 (d, 1H, J=1.8), 7.06 (s, 2H), 7.03 (d, 1H, J=2.1), 7.01 (d, 1H, J=1.5), 6.86 (d, 1H, J=1.5), 4.95 (s, 2H), 3.99 (s, 3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.79 (s, 3H), 3.37 (q, 2H, J=6.0), 3.10 (q, 2H, J=6.0), 2.98 (m, 2H), 2.74 (s, 3H), 2.72 (s, 3H), 2.34 (t, 2H, J=7.2), 1.73 (m, 2H). MALDI-TOF MS 813.6 (M+H) (812.39 calcd. for C$_{38}$H$_{48}$N$_{14}$O$_7$).

ImImPyPy(Am)Py-β-Dp (**PA7**) See **PA6** for synthesis. UV λmax (H20) (ε) 311 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 10.35 (s, 1H), 9.99 (s, 1H), 9.91 (s, 1H), 9.73 (s, 1H), 9.30 (bs, 1H), 8.03 (m, 2H), 7.56 (s, 1H), 7.45 (s, 1H,), 7.24 (s, 1H), 7.22 (s, 1H), 7.21 (s, 1H), 7.15 (s, 1H), 7.12 (s, 1H), 7.09 (s, 1H), 7.07 (s, 1H), 6.99 (m, 1H), 6.83 (s, 1H), 4.93 (s, 2H), 3.99 (s, 6H), 3.84 (s, 3H), 3.78 (s, 3H), 3.36 (q, 2H, J=6.6), 3.09 (q, 2H, J=6.0), 2.98 (m, 2H), 2.73 (s, 3H), 2.71 (s, 3H), 2.33 (t, 2H, J=6.6), 1.73 (m, 2H). MALDI-TOF MS 814.5 (M+H) (813.41 calcd. for C$_{37}$H$_{47}$N$_{15}$O$_7$).

ImPyDsPyPy-β-Dp (**PA8**) UV λmax (H20) (ε) 316 nm (43,450). ¹H NMR (DMSO-d₆) δ 11.21 (s, 1H), 10.48 (s, 1H), 9.95 (s, 1H), 9.93 (s, 1H), 9.91 (s, 1H), 9.34 (bs, 1H), 8.04 (m, 2H), 7.41 (s, 1H), 7.29 (d, 1H, J=2.1), 7.24 (d, 1H, J=1.8), 7.16 (m, 4H), 7.07 (d, 1H, J=3.0), 7.00 (d, 1H, J=1.8), 6.86 (d, 1H, J=1.8), 3.99 (s, 3H), 3.85 (s, 3H), 3.84 (s, 3H), 3.80 (s, 3H), 3.37 (q, 2H, J=5.7), 3.10 (q, 2H, J=6.3), 3.00 (m, 2H), 2.74 (s, 3H), 2.72 (s, 3H), 2.34 (t, 2H, J=7.8), 1.74 (m, 2H). MALDI-TOF MS 756.5 (M+H) (755.37 calcd. for $C_{36}H_{45}N_{13}O_6$).

ImImDsPyPy-β-Dp (**PA9**) UV λmax (H20) (ε) 316 nm (43,450).

ImDsPyPyPy-β-Dp (**PA10**) UV λmax (H20) (ε) 316 nm (43,450). ¹H NMR (DMSO-d₆) δ 11.30 (s, 1H), 10.46 (s, 1H), 9.95 (s, 2H), 9.89 (s, 1H), 9.27 (bs, 1H), 8.03 (m, 2H), 7.40 (s, 1H), 7.27 (d, 1H, J=2.1), 7.24 (d, 1H, J=1.8), 7.21 (m, 2H), 7.15 (d, 1H, J=1.8), 7.06 (s, 2H), 7.02 (d, 1H, J=1.8), 6.87 (d, 1H, J=1.8), 3.99 (s, 3H), 3.85 (s, 3H), 3.83 (s, 3H), 3.79 (s, 3H), 3.36 (q, 2H, J=6.0), 3.10 (q, 2H, J=6.3), 2.95 (m, 2H), 2.74 (s, 3H), 2.72 (s, 3H), 2.34 (t, 2H, J=7.2), 1.73 (m, 2H). MALDI-TOF MS 756.5 (M+H) (755.37 calcd. for $C_{36}H_{45}N_{13}O_6$).

ImImPyDsPy-β-Dp (**PA11**) UV λmax (H20) (ε) 316 nm (43,450). ¹H NMR (DMSO-d₆) δ 11.18 (s, 1H), 10.34 (s, 1H), 9.95 (s, 1H), 9.89 (s, 1H), 9.731 (s, 1H), 9.24 (bs, 1H), 8.04 (m, 2H), 7.57 (s, 1H), 7.46 (s, 1H), 7.27 (s, 1H), 7.18 (s, 1H), 7.14 (m, 3H), 7.07 (s, 1H), 6.80 (d, 1H, J=1.8), 3.99 (s, 6H), 3.84 (s, 3H), 3.79 (s, 3H), 3.37 (q, 2H, J=5.7), 3.10

(q, 2H, J=6.3), 3.00 (m, 2H), 2.74 (s, 3H), 2.72 (s, 3H), 2.34 (t, 2H, J=7.8), 1.72 (m, 2H). MALDI-TOF MS 757.4 (M+H) (756.37 calcd. for $C_{36}H_{45}N_{13}O_6$).

ImPyPy(2Am)PyPy-β-Dp (**P12**) UV λmax (H20) (ε) 311 nm (43,450). MALDI-TOF MS 827.6 (M+H) (826.41 calcd. for $C_{39}H_{51}N_{14}O_6$).

ImPyPy(2P)PyPy-β-Dp (**PA13**) UV λmax (H20) (ε) 311 nm (43,450). ^{1}H NMR (DMSO-d_6) δ 10.49 (s, 1H), 10.14 (s, 1H), 9.90 (s, 1H), 9.71 (s, 1H), 9.40 (bs, 1H), 8.04 (m, 2H), 7.93 (bs, 3H), 7.58 (d, 1H, J=5.7), 7.45 (s, 2H), 7.26 (s, 1H), 7.25 (s, 1H), 7.15 (s, 1H), 7.06 (d, 1H, J=2.7), 6.98 (s, 1H), 6.85 (s, 1H), 4.44 (t, 2H, J=6.0), 3.99 (s, 6H), 3.83 (s, 3H), 3.78 (s, 3H), 3.35 (q, 2H, J=5.4), 3.22 (m, 2H), 3.09 (q, 2H, J=6.0), 2.98 (m, 2H), 2.72 (s, 3H), 2.71 (s, 3H), 2.32 (t, 2H, J=5.4), 1.72 (m, 2H). MALDI-TOF MS 800.71 (M+H) (799.41 calcd. for $C_{37}H_{49}N_{15}O_6$).

ImPyPy(3P)PyPy-β-Dp (**PA14**) UV λmax (H20) (ε) 311 nm (43,450). ^{1}H NMR (DMSO-d_6) δ 10.48 (s, 1H), 10.04 (s, 2H), 9.90 (s, 1H), 9.30 (bs, 1H), 8.05 (m, 2H), 7.71 (bs, 3H), 7.40 (s, 2H), 7.29 (s, 1H), 7.26 (s, 1H), 7.18 (d, 1H, J=1.5), 7.15 (d, 1H, J=1.5), 7.09 (s, 1H), 7.05 (s, 1H), 6.99 (s, 1H), 6.87 (s, 1H), 4.37 (t, 2H, J=5.4), 3.98 (s, 3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.79 (s, 3H), 3.36 (q, 2H, J=5.9), 3.10 (q, 2H, J=6.0), 3.00 (m, 2H), 2.77 (t, 2H, J=4.2), 2.74 (s, 3H), 2.72 (s, 3H), 2.34 (t, 2H, J=6.6), 1.99 (m, 2H), 1.72 (m, 2H).MALDI-TOF MS 813.5 (M+H) (812.41 calcd. for $C_{39}H_{52}N_{14}O_6$).

ImImPy(3P)PyPy-β-Dp (**PA15**) UV λmax (H20) (ε) 311 nm (43,450). MALDI-TOF MS 814.45 (M+H) (813.41 calcd. for $C_{38}H_{51}N_{15}O_6$).

ImPyPy(3G)PyPy-β-Dp (**PA16**) **PA14** (1.6 μmol) was dissolved in 0.12 mL CH_3CN. 0.06 mL of a 0.387 M (16 μmol) solution of N,N'-Boc-guanidyl pyrazole was added and the reaction heated to 75 ºC for 2 h. The solvent was removed *in vacuo* and the residue taken up in 0.2 mL 50% TFA:DCM. After 2 h at room temperature, the solvent was evaporated and the residue taken up in 0.1% TFA. **PA16** was purified by reversed phase HPLC. UV λmax (H20) (ε) 311 nm (43,450). ^{1}H NMR (DMSO-d_6) δ 10.49 (s, 1H), 9.98 (s, 2H), 9.93 (s, 1H), 9.34 (bs, 1H), 8.05 (m, 2H), 7.41 (s, 1H), 7.30 (s, 1H), 7.297 (s, 1H), 7.28 (s, 1H), 7.26 (m, 2H), 7.24 (s, 1H), 7.21 (s, 1H), 7.18 (t, 2H, J=1.5), 7.11 (d, 1H, J=1.5), 7.08 (d, 1H, J=1.5), 7.07 (d, 1H, J=1.5), 4.51 (t, 2H, J=6.0), 4.00 (s, 3H), 3.86 (s, 3H), 3.85 (s, 3H), 3.81 (s, 3H), 3.37 (m, 4H), 3.11 (q, 2H, J=6.0), 3.00 (m, 4H), 2.75 (s, 3H), 2.73 (s, 3H), 2.35 (t, 2H, J=7.2), 1.73 (m, 2H) MALDI-TOF MS 855.66 (M+H) (854.45 calcd. for $C_{40}H_{54}N_{16}O_6$).

ImPyPy(5F)PyPy-β-Dp (**PA17**) UV λmax (H20) (ε) 314 nm (43,450). ^{1}H NMR (DMSO-d_6) δ 10.47 (s, 1H), 10.05 (s, 1H), 9.98 (s, 1H), 9.91 (s, 1H), 9.46 (bs, 1H), 8.04 (m, 2H), 7.39 (s, 1H), 7.26 (s, 1H), 7.23 (d, 1H, J=1.8), 7.17 (s, 1H), 7.15 (s, 2H), 7.04 (s, 1H), 7.03 (d, 1H, J=1.8), 7.01 (d, 1H, J=1.9), 6.87 (d, 1H, J=1.8), 5.66 (s, 2H), 3.97 (s, 3H), 3.83 (s, 3H), 3.81 (s, 3H), 3.78 (s, 3H), 3.33 (q, 2H, J=5.4), 3.09 (q, 2H, J=6.0), 2.97 (m, 2H), 2.72 (s, 3H), 2.71 (s, 3H), 2.33 (t, 2H, J=6.6), 1.72 (m, 2H). MALDI-TOF MS 936.4 (M+H) (935.37 calcd. for $C_{43}H_{47}F_5N_{13}O_6$).

ImImPy(Bz)PyPy-β-Dp (**PA19**) UV λmax (H20) (ε) 312 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 10.41 (s, 1H), 10.03 (s, 1H), 9.87 (s, 1H), 9.72 (s, 1H), 9.29 (bs, 1H), 8.02 (m, 2H), 7.56 (d, 1H, J=2.4), 7.44 (d, 1H, J=1.5), 7.42 (s, 1H), 7.28 (m, 1H), 7.26 (d, 1H, J=1.5), 7.19 (m, 1H), 7.16 (s, 1H), 7.14 (s, 1H), 7.13 (s, 1H), 7.12 (s, 1H), 7.10 (s, 1H), 7.06 (d, 1H, J=2.7), 7.02 (s, 1H), 6.85 (s, 1H), 5.59 (s, 2H), 3.98 (s, 6H), 3.80 (s, 3H), 3.77 (s, 3H), 3.35 (q, 2H, J=6.0), 3.08 (q, 2H, J=6.0), 2.98 (m, 2H), 2.72 (s, 3H), 2.70 (s, 3H), 2.32 (t, 2H, J=6.6), 1.71 (m, 2H). MALDI-TOF MS 847.28 (M+H) (846.41 calcd. for C$_{42}$H$_{51}$N$_{14}$O$_6$).

ImPyPy(2Bz)PyPy-β-Dp (**PA20**) UV λmax (H20) (ε) 312 nm (43,450).

ImImPy(2Bz)PyPy-β-Dp (**PA21**) UV λmax (H20) (ε) 312 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 10.33 (s, 1H), 9.99 (s, 1H), 9.90 (s, 1H), 9.73 (s, 1H), 9.36 (bs, 1H), 8.03 (m, 2H), 7.56 (s, 1H), 7.45 (s, 1H), 7.28 (s, 2H), 7.24 (s, 1H), 7.23 (s, 2H), 7.21 (m, 1H), 7.19 (s, 1H), 7.15 (m, 2H), 7.06 (m, 2H), 6.86 (s, 1H), 4.49 (t, 2H, J=6.0), 3.99 (s, 6H), 3.84 (s, 3H), 3.78 (s, 3H), 3.36 (q, 2H, J=6.0), 3.09 (q, 2H, J=6.0), 2.98 (m, 4H), 2.72 (s, 3H), 2.71 (s, 3H), 2.33 (t, 2H, J=6.6), 1.72 (m, 2H) MALDI-TOF MS 861.16 (M+H) (860.43 calcd. for C$_{43}$H$_{52}$N$_{14}$O$_6$).

ImPyPy(In)PyPy-β-Dp (**PA22**) UV λmax (H20) (ε) 314 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 10.47 (s, 1H), 9.86 (s, 2H), 9.73 (s, 1H), 9.19 (bs, 1H), 8.03 (m, 2H), 7.39 (s, 1H), 7.31 (s, 1H), 7.21 (s, 1H), 7.19 (s, 1H), 7.16 (s, 1H), 7.14 (s, 1H), 7.04 (s, 1H), 7.02 (s, 1H), 6.94 (m, 1H), 6.85 (s, 1H), 6.60 (m, 1H), 6.55 (m, 1H), 5.32 (d, 1H), 4.54 (s,

2H), 4.48 (bs, 1H), 3.98 (s, 3H), 3.87 (s, 3H), 3.81 (s, 3H), 3.78 (s, 3H), 3.09 (q, 2H, J=6.3), 2.98 (m, 2H), 2.72 (s, 3H), 2.71 (s, 3H), 2.32 (t, 2H, J=6.6), 1.71 (m, 2H) MALDI-TOF MS 887.51 (M+H) (886.44 calcd. for $C_{45}H_{54}N_{14}O_6$).

Im(iBu)DsPyPyPy(iBu)-β-Dp (**PA23**) UV λmax (H20) (ε) 316 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 11.23 (s, 1H), 10.44 (s, 1H), 9.96 (s, 2H), 9.90 (s, 1H), 9.20 (bs, 1H), 8.03 (m, 2H), 7.41 (s, 1H), 7.30 (s, 1H), 7.24 (s, 1H), 7.22 (s, 1H), 7.18 (s, 1H), 7.17 (s, 1H), 7.06 (s, 1H), 7.05 (d, 1H, J=1.8), 7.01 (s, 1H), 6.85 (s, 1H), 4.28 (d, 2H, J=7.8), 4.06 (d, 2H, J=6.6), 3.84 (s, 3H), 3.83 (s, 3H), 3.08 (m, 2H), 2.97 (m, 2H), 2.74 (s, 3H), 2.72 (s, 3H), 2.34 (t, 2H, J=7.2), 2.07 (m, 1H), 1.92 (m, 1H), 1.76 (m, 2H), 0.83 (d, 6H, J=6.3), 0.77 (d, 6H, J=6.3) MALDI-TOF MS 840.8 (M+H) (839.46 calcd. for $C_{42}H_{57}N_{13}O_6$).

ImImPy(iBu)DsPy-β-Dp (**PA24**) UV λmax (H20) (ε) 316 nm (43,450). ^{1}H NMR (DMSO-d$_6$) δ 11.20 (s, 1H), 10.36 (s, 1H), 9.96 (s, 1H), 9.91 (s, 1H), 9.74 (s, 1H), 9.30 (bs, 1H), 8.07 (m, 2H), 7.58 (s, 1H), 7.47 (s, 1H), 7.32 (d, 1H, J=1.8), 7.20 (d, 1H, J=1.5), 7.17 (s, 1H), 7.12 (s, 2H), 7.08 (s, 1H), 6.82 (d, 1H, J=1.8), 4.13 (d, 2H, J=6.9), 4.02 (s, 3H), 4.01 (s, 3H), 3.81 (s, 3H), 3.39 (q, 2H, J=6.6), 3.11 (q, 2H, J=5.7), 3.00 (m, 2H), 2.75 (s, 3H), 2.73 (s, 3H), 2.35 (t, 2H, J=7.2), 1.99 (m, 1H), 1.74 (m, 2H), 0.81, (d, 6H, J=6.6). MALDI-TOF MS 799.5 (M+H) (798.41 calcd. for $C_{38}H_{50}N_{14}O_6$).

Preparation of 3' ^{32}P-end Labeled Restriction Fragments.

Plasmid pDHN2 was cut using *PvuII* and *EcoRI* to yield a 250 bp restriction fragment containing the binding sites 5'-caTGGTACAt-3', 5'-caTGGTCCAt-3', and 5'-

caTGTTACAt-3'. The sticky ends of the fragment were filled in using Sequenase, [α-^{32}P]-deoxyadenosine-5'-triphosphate, and [α-^{32}P]-thymidine-5'-triphosphate. The labeled fragment was then purified by nondenaturing gel electrophoresis. A and G sequencing were performed as previously described.[17]

Quantitative DNase I Footprinting Titrations.

Polyamide/DNA equilibrations and DNase I footprinting were performed, and equilibrium association constants determined, as previously described for all homodimer experiments. In heterodimer experiments, the polyamide concentration reported is the concentration of each individual polyamide. Thus, a 100 nM equilibration will contain each of the two polyamides at 100 nM for a total polyamide concentration of 200 nM. All other experimental parameters remain unchanged from the homodimer protocol.[9]

References.

1. Stryer, L., *Biochemistry*. W. H. Freeman and Co.: New York, 1995.

2. Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A., *Science*, **2000,** *289*, 905–920.

3. Gallivan, J. P.; Dougherty, D. A., *Proceedings of the National Academy of Sciences of the United States of America*, **1999,** *96*, 9459–9464.

4. Gallivan, J. P.; Dougherty, D. A., *Journal of the American Chemical Society*, **2000,** *122*, 870–874.

5. Ma, J. C.; Dougherty, D. A., *Chemical Reviews*, **1997,** *97*, 1303–1324.

6. West, A. P.; Mecozzi, S.; Dougherty, D. A., *Journal of Physical Organic Chemistry*, **1997,** *10*, 347–350.

7. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C., *Nature Struct. Biol.*, **1998,** *5*, 104.

8. Trauger, J. W., *Ph. D. Thesis*, California Institute of Technology, Pasadena, **1998**, 77.

9. Trauger, J. W.; Dervan, P. B., Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. In *Drug-Nucleic Acid Interactions*, 2001; Vol. 340, pp 450–466.

10. Pardo, C.; et al., *OPPI Briefs*, **2000,** *4*, 385–390.

11. Rucker, V. C.; Fostier, S.; Melander, C.; Dervan, P. B., *Journal of the American Chemical Society*, **2003,** *125*, 1195–1202.

12. Bremer, R. E.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2000,** *8*, 1947–1955.

13. Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6141–6146.

14. Nguyen, D. H., *Ph. D. Thesis*, **2001**.

15. Doherty, E. A.; Doudna, J. A., *Annual Review of Biochemistry*, **2000,** *69*, 597–615.

16. Pandolfi, P. P., *Oncogene*, **2001,** *20*, 3116–3127.

17. Maxam, A. M.; Gilbert, W. S., *Methods in Enzymology*, **1980,** *65*, 499.

Appendix I

DNA Pulldown: Capture Purification of DNA Fragments with Hairpin Polyamide-Biotin Conjugates

Initial pulldown experiments using fluorescently labeled oligonucleotides were done in collaboration with postdoctoral scholar Alex Heckel.

Abstract

The purification and isolation of fragments of genomic DNA would be of great use to the biochemical field. We report herein efforts towards the use of DNA-binding polyamide-biotin conjugates to isolate targeted fragments of DNA from complex mixtures using streptavidin-coated magnetic beads. Experiments done in the presence of 2 kB of DNA indicate that the polyamide specificity remains an important hurdle to moving forward.

Introduction.

Chromatin Immunoprecipitation (ChIP) has emerged as a powerful technique for identifying protein-protein and protein-DNA interactions.[1-5] In this protocol, biological DNA-protein macrostructures are crosslinked inside cells with formaldehyde. The cells are then lysed and the DNA sheared into approximately 500 base pair fragments. An antibody targeted towards a protein that is known to bind to the sequence of interest is then used to purify that fragment (and all proteins crosslinked to it). The covalent crosslinks are then reversed, and the individual components of the complex purified by gel electrophoresis and analyzed by mass spectrometry. Using this technique, all elements of a promoter of a gene of interest can be identified.

CHiP does, however, suffer from several disadvantages. First, *a priori* knowledge of at least one of the proteins bound to the sequence of interest is a prerequisite for antibody generation. Second, antibody generation is itself a tedious process that does not always yield useful binders.

Over the past ten years, researchers in the Dervan lab have developed programmable small molecules capable of recognizing predetermined sequences of DNA with high affinity and specificity.[6-8] Furthermore, these polyamides, which bind in the minor groove of the DNA duplex, have been shown to bind DNA in the presence of nearby proteins, proteins bound to the same site but in the major groove, and even to DNA packaged on chromatin.[6, 9] Because of their ability to bind DNA (and even protein-DNA complexes) with affinities and specificities rivaling natural DNA-binding proteins, we explore the use polyamides as molecular tools for the isolation of specific DNA fragments from complex mixtures.[10]

Because of recent genome sequencing efforts, the base pair content of almost every human gene and promoter is known.[11] The identities of proteins bound to a given promoter at any given time are not as well understood. Because of precise sequence information, polyamides can be designed to target a promoter of interest. DNA-protein complexes can be crosslinked inside cells using formaldehyde, and the DNA sheared into 500 base pair fragments. By conjugating a chemical handle to a polyamide that is capable of binding the sequence of interest with high affinity and specificity, the sequence of interest, and all of its associated proteins, can be purified, or pulled down, from the mixture of oligonucleotide fragments.

The use of DNA-binding polyamides for this type of analysis has several advantages over CHiP. First, polyamides can be easily designed and synthesized to target almost any sequence of DNA. Second, polyamides become useful for studying promoters for which there are no known protein binders. We report herein progress towards the use of polyamides as "pulldown" reagents for DNA capture.

DNA Pulldown with Biotin-Polyamide Conjugates.

Many different pulldown architectures can be envisioned. For example, polyamides could be conjugated to a solid support, such as polystyrene or polyacrylamide beads. Alternatively, polyamides may be functionalized with a chemical handle that can later be recognized by a functionalized solid support. This method has the advantage that the polyamides can equilibrate with the DNA in solution, rather than on solid support, which may impede the equilibration process. We chose to use polyamide-biotin
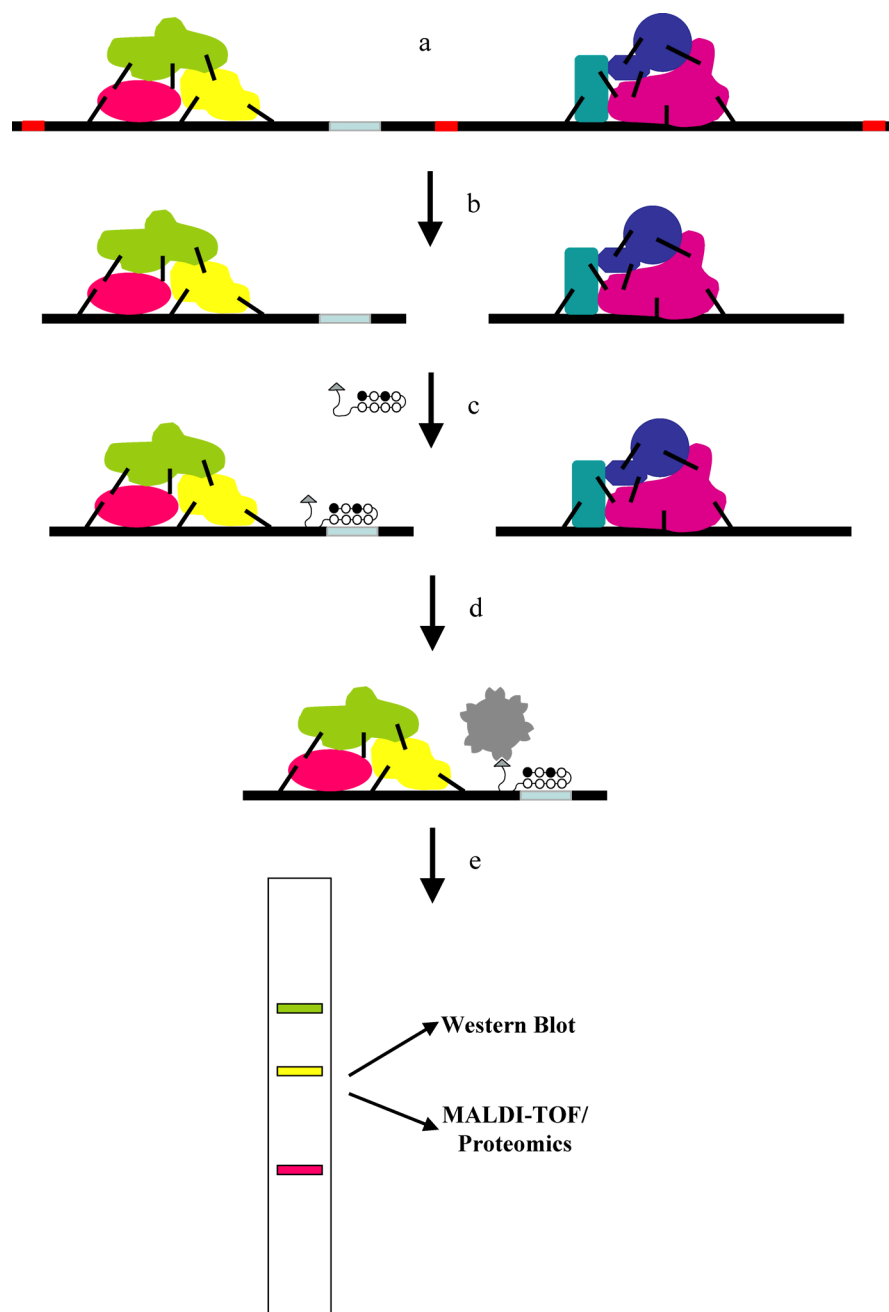
Scheme I.1. Schematic procedure for isolating a locus of interest (left side) from genomic DNA. Genomic DNA is crosslinked to associated proteins using formaldehyde. (a). DNA is then fragmented (at red squares) either by shearing or restriction digest. (b). A polyamide-biotin conjugate designed for the locus of interest (blue square = binding site) is then incubated with the DNA pieces. (c). The fragments are then incubated with streptavidin-coated beads. After washing, only the locus of interest remains bound via the polyamide-biotin:streptavidin interaction. (d). The fragment of interest is then eluted from the beads, the formaldehyde crosslinks reversed, and the proteins analyzed by gel electrophoresis and mass spectrometry.

conjugates. The DNA-polyamide complexes can then be isolated from DNA fragments that do not bind the polyamide by incubation with streptavidin-coated beads. The biotin-streptavidin interaction is one of the strongest non-covalent interactions known. For these studies, streptavidin-coated magnetic beads are used (commercially available from Dynal, www.dynal.no). Thus, after immobilization of the polyamide:DNA complex on the streptavidin-coated beads, the supernatant can be facilely removed by using a magnet to collect and trap the bead complexes. Scheme I.1 illustrates a hypothetical pulldown experiment with polyamide-biotin conjugates.
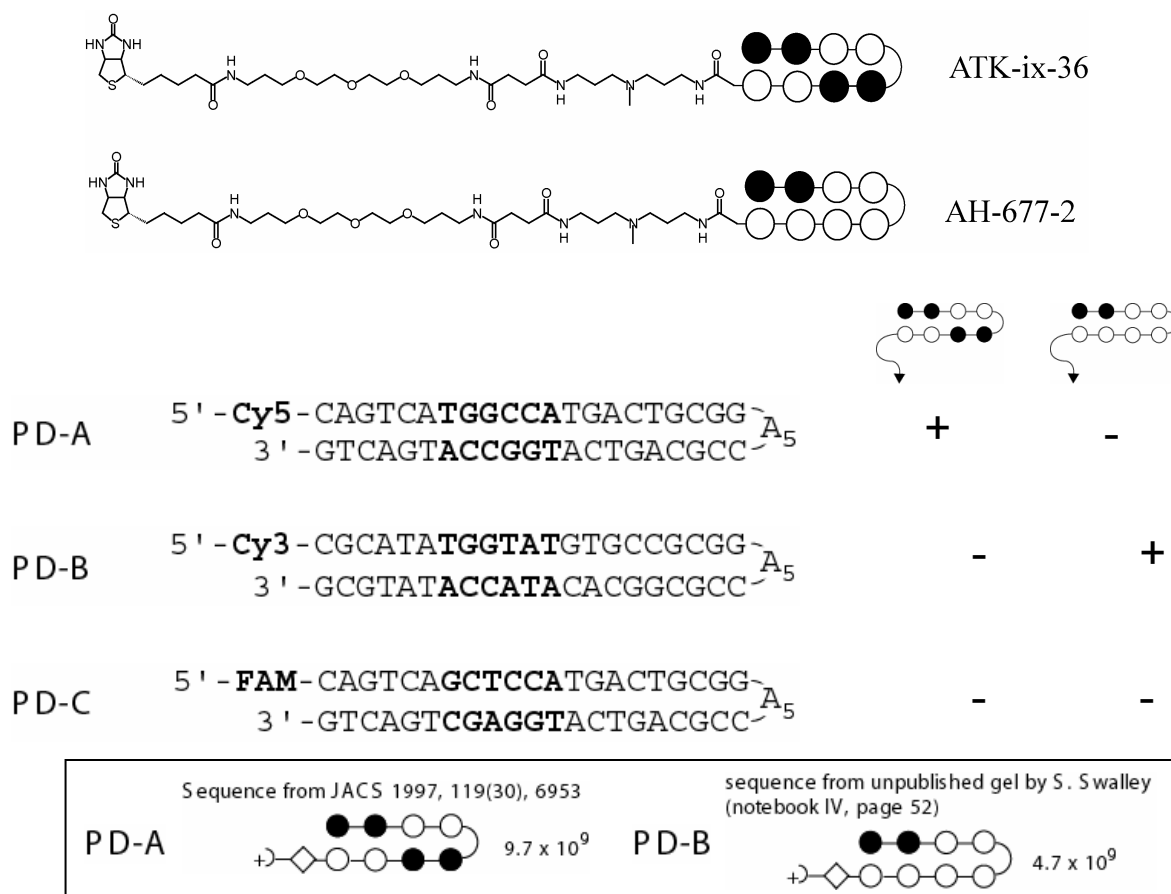
Figure I.1. Top: Compounds used in initial pulldown experiments. Black and white circles represent imidazole and pyrrole carboxamides, respectively. Bottom: Fluorescent duplexes used for pulldown studies. **PD-A** contains a match site for **ATK-ix-36**, **PD-B** contains a match site for **AH-677-2**, **PD-C** is a negative control. Boxed are measured binding affinities for the compounds (non-conjugated) used in this study. Plus represents match binding site, minus represents mismatch binding site.

The focus of initial experiments is to determine whether polyamide-biotin conjugates can isolate specific fragments of DNA from a mixture containing other fragments with formal mismatch binding sites. Polyamides **ATK-ix-36** and **AH-677-2** (synthesized by Alex Heckel) were synthesized by reacting polyamide amines with the commercially available biotinylation reagent TFP-PEO-Biotin (Pierce) (Figure I.1). Additionally, three duplex DNA oligonucleotides were synthesized containing match binding sites for **ATK-ix-36** (**PD-A**), **AH-677-2** (**PD-B**), and a negative control double mismatch site (**PD-C**) (Figure I.1). The three oligonucleotides were tagged with fluorophores with distinct excitation and emission spectra. This allows for the separation of each component's signal from a solution containing all three dyes, thereby allowing the amount of each oligo to be quantified.

Initial experiments to explore the specificity and yields of polyamide pulldown were then performed. Equilibrations and pulldown reactions are performed in 50 μL reaction volumes of TKMC/T20 (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl$_2$, 5 mM CaCl$_2$, pH 7.0, 0.1% v/v Tween 20). An equimolar mixture of the three duplex oligonucleotides was incubated with a biotinylated polyamide (either **ATK-ix-36** or **AH-677-2**) for 15 minutes at room temperature. This equilibration time was shown to be sufficient to establish maximum binding selectivity.[12] After equilibration, streptavidin-coated magnetic beads were added, and the suspension shaken (600 rpm) for 15 minutes at room temperature. Again, this time was shown to be sufficient to maximally absorb all biotin complexes to the beads.[12] The pulldown capacity of the magnetic beads was experimentally determined to be 100–500 pmol biotin per μg bead.[12] After the incubation, the beads were concentrated and immobilized with a magnet, and the
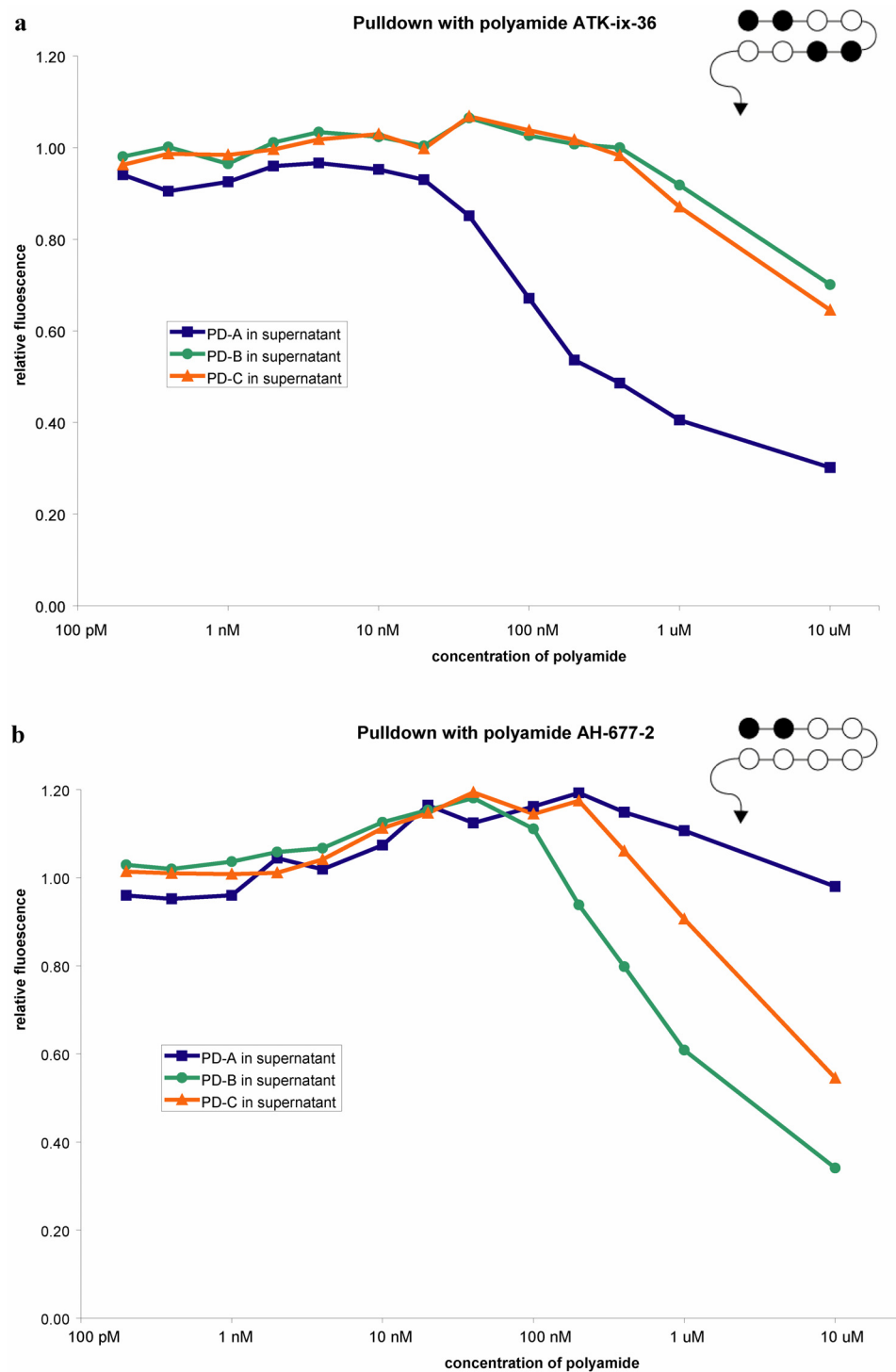
Figure I.2. a. Pulldown results with **ATK-ix-36**. **PD-A** is the match oligo, **PD-B** and **PD-C** are mismatch oligos. b. Pulldown results with **AH-677-2**. **PD-B** is the match oligo, **PD-A** and **PD-C** are mismatch oligos.

supernatant removed. The beads were then washed once with a fresh aliquot of buffer. Control vials in which no magnetic beads were added and vials in which no polyamide-biotin conjugate was added were run side-by-side with the pulldown reactions. 20 μL of each solution (control, supernatant, wash) was placed into wells on a 96-well plate. The plate was placed on a Typhoon phosphorimager and fluorescence was detected in each well. Using the ImageQuant software package, pulldown yields and efficiencies were calculated according to the following equation.

Pulldown yield = 1 / [((fluorescence in supernatant + wash)-2 x background)/(fluorescence in control-background)] x 100

Because polyamide binding and sequence specificity are dependent upon concentration, initial pulldown experiments were carried out over 5 orders of magnitude in concentration to determine the optimal concentration for both pulldown yield and selectivity for match oligo. Data for **ATK-ix-36** and **AH-677-2** are illustrated in figure I.2. As shown, at 300 nM, **ATK-ix-36** is able to capture 55% of **PD-A**, the duplex containing its match binding site, while leaving more than 95% of the two mismatch oligos (**PD-B** and **PD-C**) in solution. Even at 1 μM, **ATK-ix-36** is relatively specific, capturing 60% of **PD-A**, 4% of **PD-B**, and 3% of **PD-C**. At 300 nM, **AH-677-2** is less efficient, capturing only 20% of its match oligo **PD-B**. **AH-677-2** is also less specific, pulling down 7% of **PD-C** at 300 nM. At 1 μM, this polyamide captures 38% of match oligo **PD-B**, 25% of **PD-C**, and 7% of **PD-A**. While yields at 100 nM are not maximal, higher concentrations of polyamide-biotin conjugates led to significant non-specific pulldown. Thus, for future experiments, concentrations of 100 nM were used.

Release of Captured DNA.

Because our eventual goal is to use this technology to purify interesting DNA fragments, it is necessary to be able to recover the DNA of interest from the streptavidin-coated beads. Furthermore, our interest in this project is to purify large pieces of DNA from fragmented genomes. On the genomic scale, a typical 8-ring hairpin polyamide binding site can be found approximately one million times. Thus, we anticipate that in order to purify sequences from genomic DNA, multiple pulldowns will have to be performed. That is, in a genome that has been fragmented into approximately 500 base pair pieces, an 8-ring hairpin polyamide-biotin conjugate may pull out dozens of fragments where match sites are present and accessible. In order to fully isolate a single fragment, a second (and possibly a third, etc.) pulldown with a different polyamide will be done. Now, because the second pool contains fewer fragments, the probability of multiple fragments having an additional common sequence is reduced, and the fragment of interest can be further purified (Scheme I.2).

Initial attempts to disrupt either the biotin-streptavidin interaction or the polyamide-DNA interaction to release the captured DNA were unsuccessful. These include treatment with 3M NaOAc, pH 5.2, 37 °C, 10 h; 4M NaCl, 37 °C, 14 h; 50% formamide$_{(aq)}$, 90 °C, 10 min; Chloroquine (intercalator), rt, 2 h; 1M guanidinium chloride, 37 °C, 1 h; or N-methylamino dipropylamine in water, 37 °C, 14 h.[12] While heating to 90 °C did release ~30% of the captured DNA, this low yield coupled with incomplete pulldown led us to explore other options. In a final attempt, a solution of excess biotin was able to release ~90% of the captured DNA from beads. However, this leaves the polyamide-biotin conjugate bound
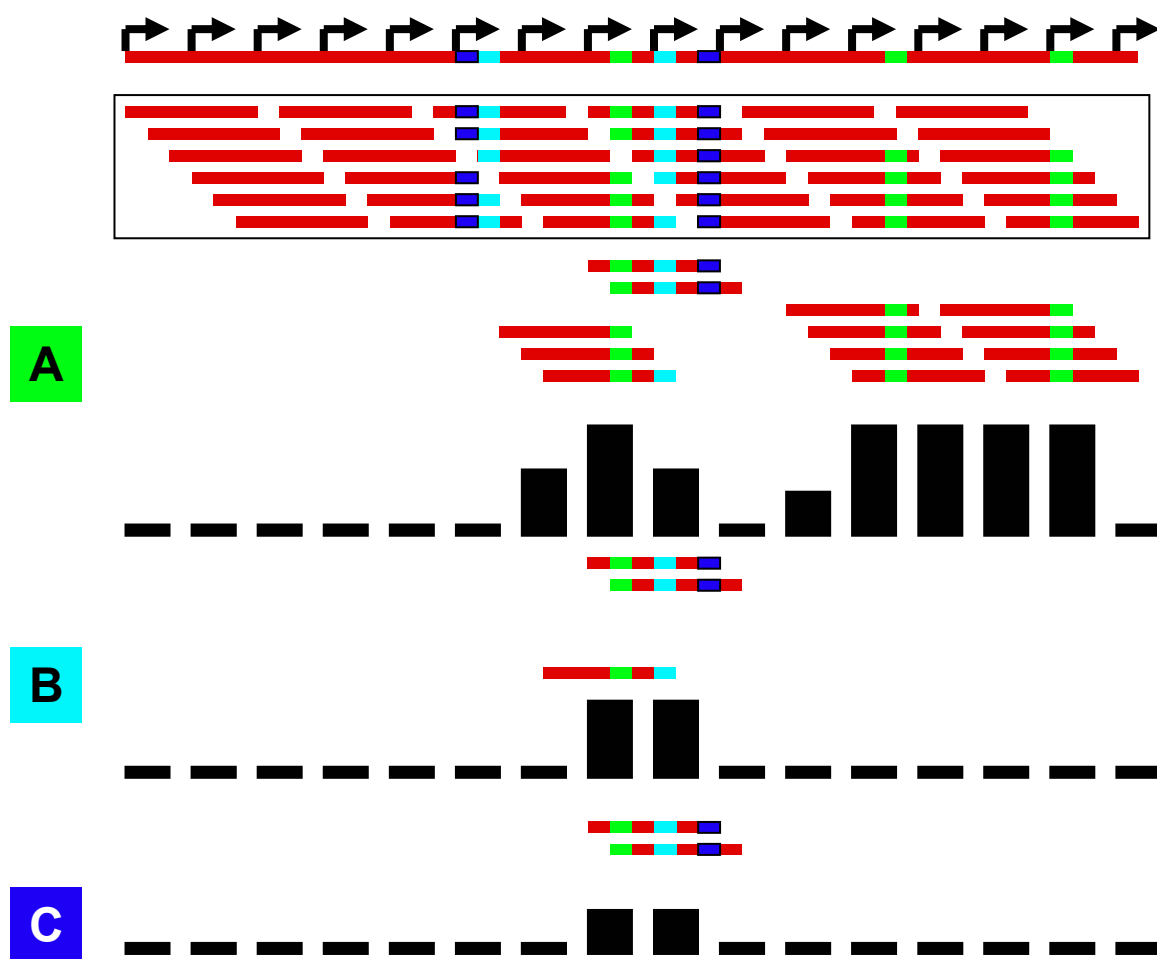
Scheme I.2. Schematic representation of how multiple pulldowns with different polyamides can be used to purify single fragments. Genomic DNA is fragmented (Boxed). Polyamide A is used to pulldown all fragments with match sites (green box). The fragments of interest (pieces with 3 binding sites) are now enriched, but impurities (all fragments on the right) are still present. Subsequent pulldown with either polyamide B or C results in the isolation of only the fragments of interest.

to the DNA, and attempts to extract the polyamide from the DNA were unsuccessful.

Desthiobiotin and iminobiotin are two commercially available biotin analogs that have reduced affinity for streptavidin. Desthiobiotin binds streptavidin with 100,000-fold lower affinity than biotin. Iminobiotin is a switchable streptavidin binder, associating tightly with streptavidin at pH > 9.0, but dissociating under acidic conditions due to protonation of the imine. Analogs of **ATK-ix-36** containing desthiobiotin and

a

b
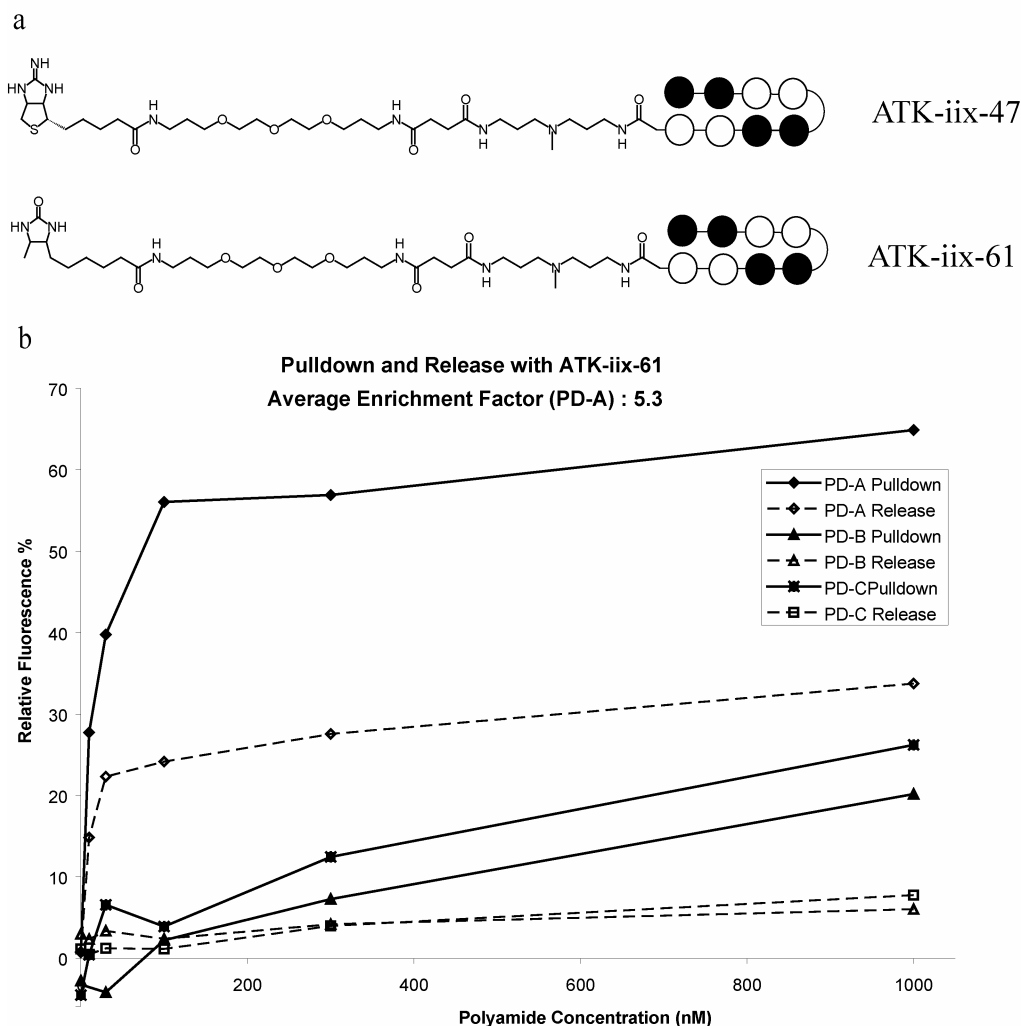
Figure I.3. a. Structures of polyamides conjugated to biotin analogs. **ATK-iix-47** was unable to capture significant amounts of DNA. b. Pulldown and release data for polyamide desthiobiotin conjugate **ATK-iix-61** (release done by incubating beads (with captured oligos) with 1 mM biotin (0.1 % Tween 20).). Enrichment factor calculated by dividing the amount **PD-A** by the amount of **PD-B** or **PD-C** at 100 nM.

iminobiotin were synthesized and tested for their ability to pulldown and release DNA (Figure I.3). Iminobiotin conjugate **ATK-iix-61** was able to pull down 35% of **PD-A** at pH 9.0. However, during the initial washing of the beads, 60% of the captured DNA was eluted from the beads. When the beads were treated with excess biotin for 1 hour, the remaining fraction of captured DNA was eluted. Desthiobiotin conjugate **ATK-iix-47** was able to capture more than 54% of the match oligo. Treatment with 1 mM biotin for 1 hour was able to displace approximately 50% of the captured oligo, resulting in a 25%

yield of **PD-A** after release. **PD-B** and **PD-C** were present in the release fraction in 4%, and 5% of their initial concentrations. Thus, **PD-A** was enriched by ~5-fold over each of the two mismatch oligos after one pulldown and release experiment.

Release with Cleavable Linkers.

Capture and release strategies with the biotin analogs described above still suffer from serious drawbacks. When the desthiobiotin-polyamide is used, upon release the conjugate is still associated with the DNA of interest. If it becomes necessary to perform
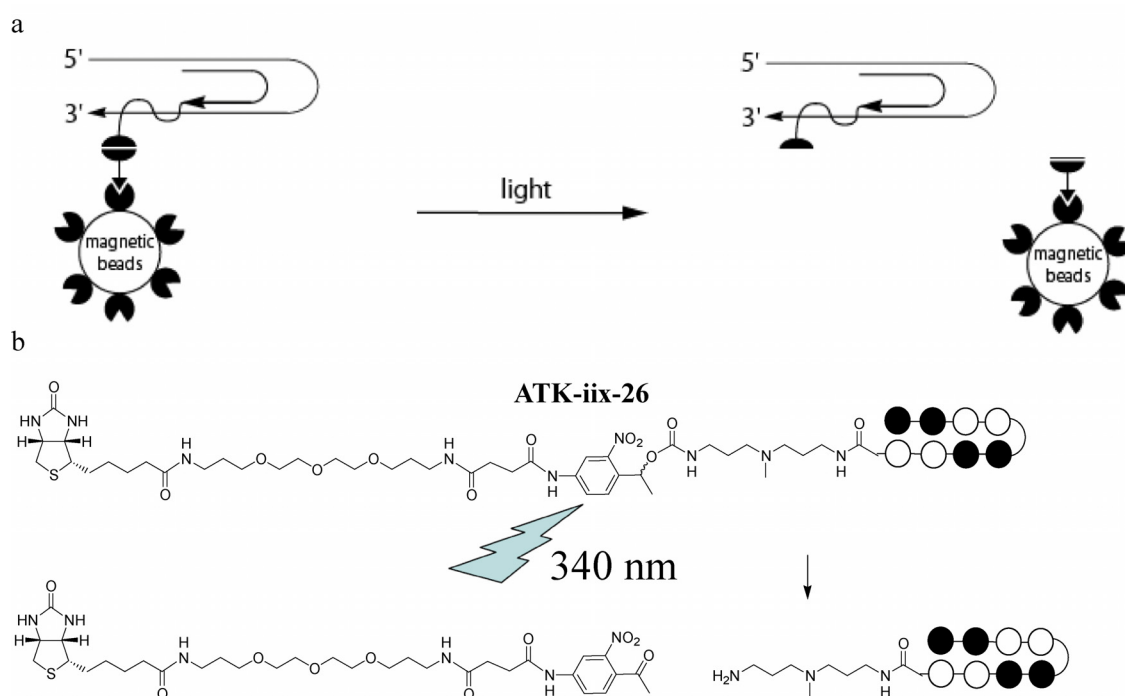
Figure I.4. a. Schematic representation of the incorporation of a photocleavable linker into a polyamide-biotin pulldown reagent. b. Chemical structures of the photocleavable polyamide-biotin conjugate **ATK-iix-26** synthesized for this project (top), and the putative structures after irradiation with 340 nm light.

multiple pulldowns to fully purify a fragment, this will lead to problems. If a second polyamide conjugate is added to the partially purified mixture isolated from a first pulldown experiment, the first pulldown polyamide is still present and no further purification can be achieved. We thus began exploring strategies to cleave the linker

between the DNA-binding portion and the streptavidin-binding portion of the capture

agents. *Ortho*-nitro benzyl esters can be cleaved by irradiation with 340 nm light (Figure

I.4). We set out to synthesize conjugate **ATK-iix-26**, which contains this photocleavable

linker (Figure I.5). Amine **5** was synthesized according to published procedures.[13] This

amine was then conjugated to PFB-PEO-Biotin (Pierce) to yield conjugate **6**. Reduction
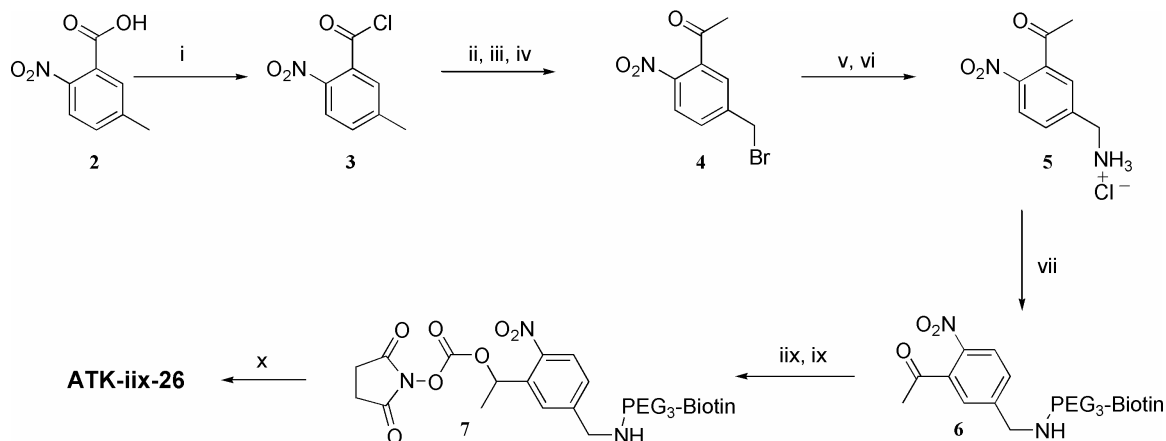
Figure I.5. Synthesis of photocleavable polyamide-biotin conjugate. i) SOCl$_2$; ii) Mg[CH(COOEt)$_2$]$_2$; iii) H$^+$/H$_2$O; iv) NBS, benzoyl peroxide; v) hexamethylene tetramine; iv) HCl, EtOH; vii) PEG$_3$-Biotin acid, DCC, HOBt; iix) NaBH$_4$; ix) Succinic anhydride, THF; x) ImImPyPy-γ-ImImPyPy-NH$_2$, DIEA, DMF.

of the ketone and subsequent NHS activation were accomplished according to published

procedures.[13] Finally, the activated biotin linker **7** was conjugated to the tail amine of

polyamide **ATK-ix-35** to yield **ATK-iix-26**.

Control experiments were done to determine photocleavage efficiency. 100 nM

solutions of **ATK-iix-26** were irradiated with a handheld UV lamp at 365 nm for 10

minutes, 1 hour, 2 hours, and 4 hours. Cleavage was monitored by reversed phase HPLC

(Figure I.6). As shown, complete cleavage was achieved after 2 hours of UV irradiation.

It should be noted that optimal photocleavage is achieved at 340 nm. The literature

reports complete cleavages in as little as 5 minutes when high-powered sources of 340

nm light are used.[13]   Because initial control experiments were successful, a pulldown

reaction was carried out using **ATK-iix-26.**
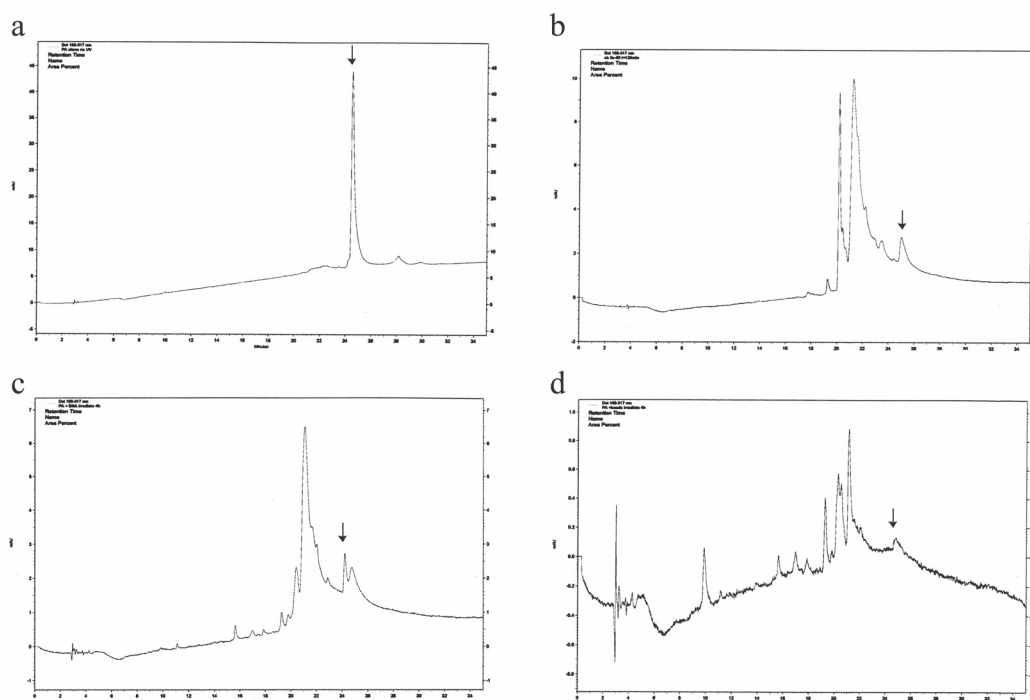
Figure I.6. Analytical HPLC traces monitoring the photocleavage of **ATK-iix-26**. a) Purified **ATK-iix-26**. In each trace, an arrow indicates the starting material. b) After 2 hours of irradiation with 365 nm light. Cleavage products are the two large peaks. c) 2 hour cleavage in the presence of 1 molar equivalent match DNA, linker is still efficiently cleaved. d) 2 hour cleavage in the presence of streptavidin-coated beads. Scale on this trace is 1/10 that of the other three traces, showing that yields for the photocleavage reaction on solid support are very low.

While 50% of the match DNA (**PD-A**) was captured by this conjugate, release by treating

with 365 nm light for 4 hours did not yield any product. Controls were done to verify

that the Cy5 dye on **PD-A** does not photobleach upon irradiation with 365 nm light. As

an additional control, **ATK-iix-26** was absorbed onto streptavidin-coated beads, and

irradiated with 365 nm light. As shown in Figure I.6, HPLC traces reveal very little

polyamide in solution, and the polyamide suffers from degradation. Thus, photocleavage

on solid support is not a feasible solution for release of captured DNA.

In a final attempt to solve the release problem, conjugate **ATK-iix-89**, which contains a chemically cleavable disulfide bridge, was synthesized (Figure I.7a). Pulldown experiments with this linker were performed as described above. Following

a

b
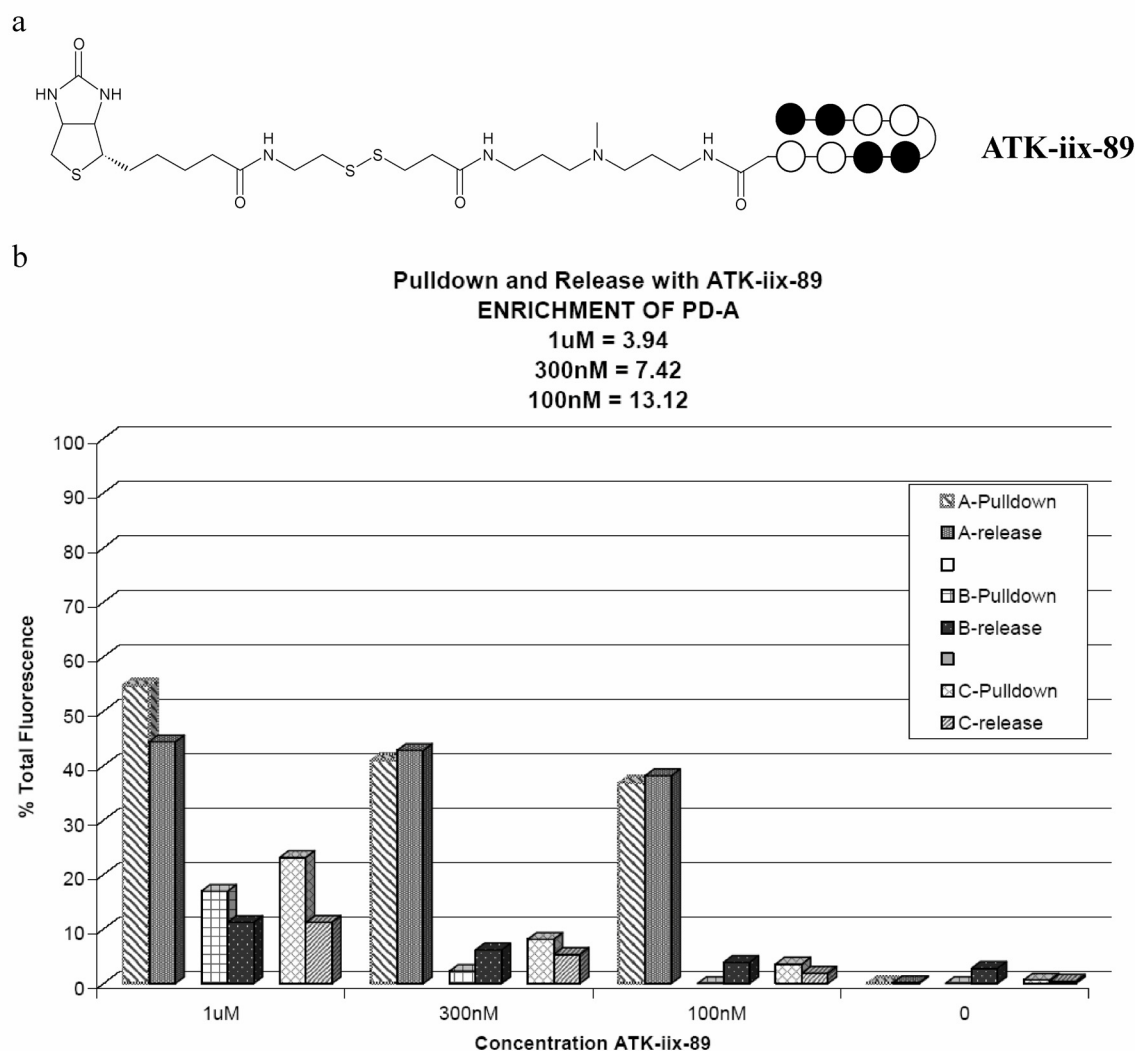
Figure I.7. a) Structure of chemically cleavable polyamide-biotin conjugate **ATK-iix-89**. b) Graphical representation of pulldown and release reaction with **ATK-iix-89**. Enrichment factors were calculated by dividing the amount of **PD-A** by the amount of **PD-B** or **PD-C**.

pulldown, release was achieved by first washing the magnetic beads with buffer, and then incubating in 50 mM dithiothreitol (DTT) for 30 minutes at 37 ºC to cleave the disulfide bond. Figure I.7b shows the results of these experiments. Conjugate **ATK-iix-89** is able

to capture almost 40% of its match DNA (**PD-A**) with less than 5% impurity at 100 nM. Furthermore, treatment with DTT is able to release 100% of the captured DNA. Thus, after a single pulldown, **PD-A** is able to be purified from **PD-B** and **PD-C** in 40% yield with a 13-fold enrichment factor.

Towards Large Fragment Pulldown

Because an efficient pulldown and release strategy had been elucidated, we were next interested in whether this strategy could be used to purify larger sequences of DNA. pUC19 is a 2686 base pair circular plasmid. A survey of restriction enzyme sites on pUC19 was made to find a restriction enzyme that would create a library of DNA fragments ranging from 100 base pairs to 1000 base pairs. MspA1 I (New England Biolabs) was identified as a restriction enzyme that fulfills the above requirements
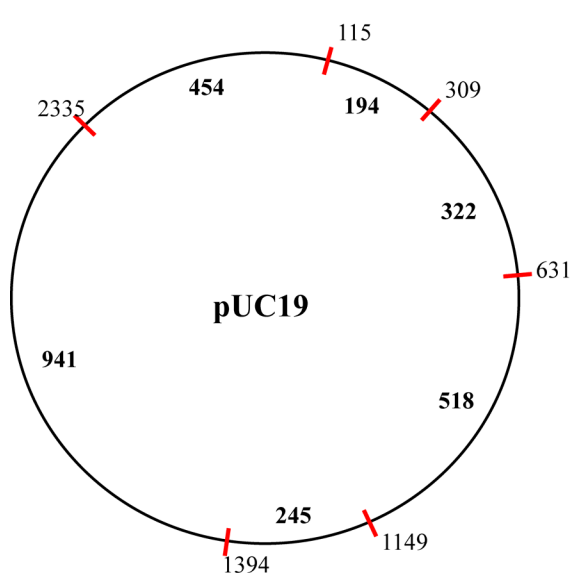
Figure I.8. Schematic representation of pUC19 digested with the restriction enzyme MspA1 I. Digestion at the indicated cut sites (red lines, with the base pair position of cleavage indicated outside the circle) yields 5 fragments ranging from 941 to 194 base pairs (fragment sizes indicated on the interior of the circle).

(Figure I.8). Cutting pUC19 with MspA1 I yields 6 fragments ranging from 194 to 941 base pairs. Additionally, the six lengths are all distinct from each other and can thus be individually visualized by non-denaturing gel electrophoresis. Pulldown experiments were then performed on a mixture of these six fragments using disulfide-bridged

polyamide-biotin conjugates.   Briefly, the mixture of oligos was incubated with

polyamide conjugate for 10 hours in TKMC/T20 buffer. Streptavidin-coated magnetic
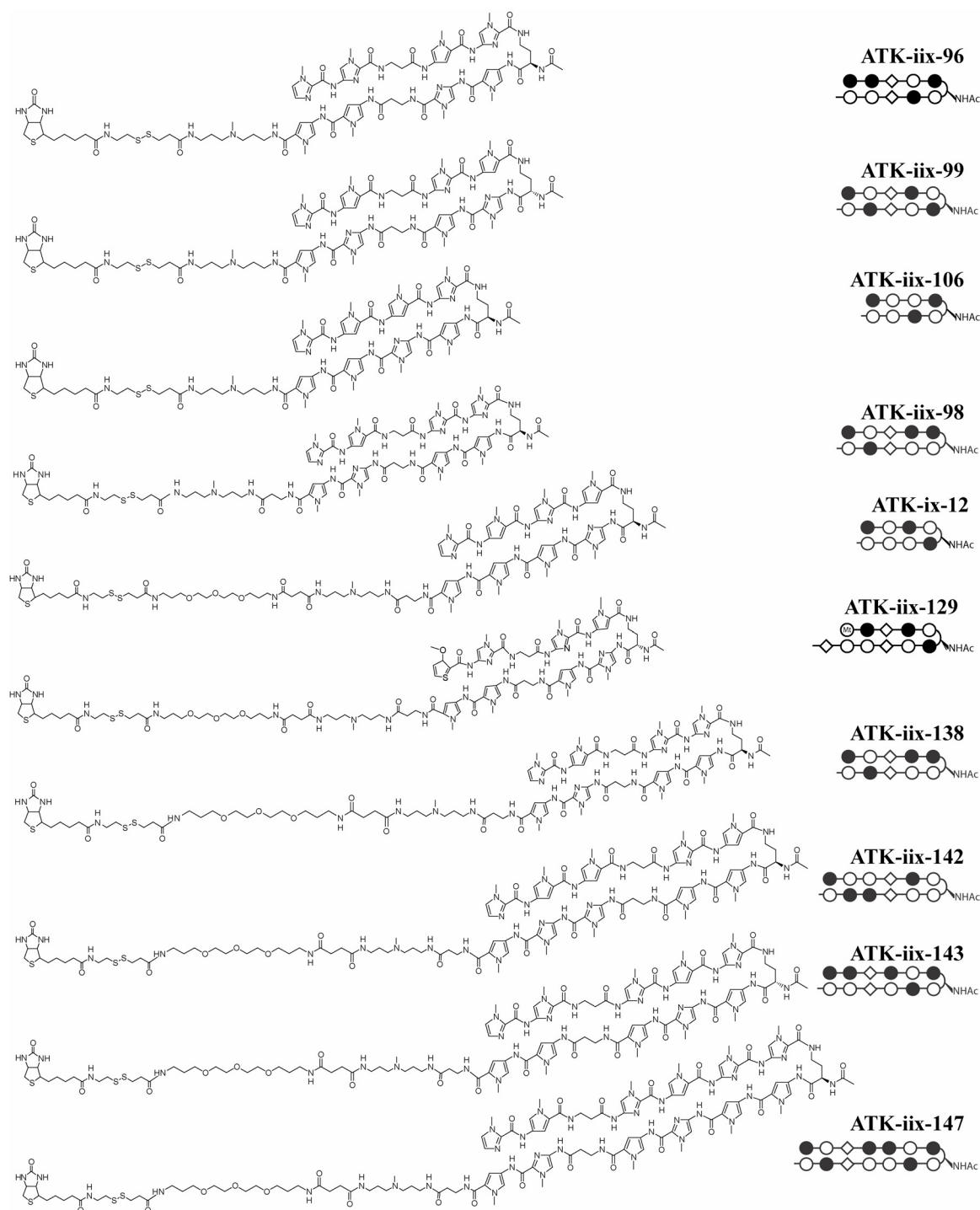
Figure I.9. Chemical structures of the hairpin polyamide-biotin conjugates synthesized for pUC19 pulldown studies (left column). Shown at right are the ball-and-stick representations of the polyamide cores for each of the conjugates. Compound names are also shown at right.

beads were then added, and the mixture shaken for 15 minutes. The beads were washed, and the DNA released by treatment with 50 mM DTT for 30 minutes. Both the supernatant (from the pulldown reaction) and the release solution were loaded onto a 5% agarose gel and run for 2 hours at 160 volts. The gels were then incubated with the DNA stain SYBR Gold (Molecular Probes) and the gels visualized with the Typhoon PhosphorImager.

A library of polyamide-SS-biotin conjugates was synthesized for these experiments (Figure I.9). The data for these experiments are illustrated in figures I.10-14. Sequence analysis of pUC19 was done to determine the locations of all match sites for each of the polyamides. For most polyamides tested, all the expected fragments were captured and released. In almost every case, the 941 base pair fragment was captured, regardless of the presence of a match site. Pulldown and release yields with these large fragments of DNA were approximately 20%.

In an effort to purify a single fragment from the complex mixture, double pulldown reactions were attempted (Figure I.15). First, DNA was captured and released using either **ATK-iix-89** or **ATK-iix-88**. The DNA in the release solution was then purified by ethanol precipitation to remove the DTT. The resulting DNA was then taken up in TKMC/T20 buffer and incubated with **ATK-iix-99** or **ATK-iix-98**, respectively. A pulldown and release was then performed on this incubation. Each reaction was then resolved by gel electrophoresis. As shown, the 518 base pair fragment is isolated in 2% overall yield in 91% purity.

Figure I.10. Pulldown experiments with **ATK-iix-89** and **ATK-iix-99**. Shown at left is a schematic of pUC19 with cut sites (red boxes) and polyamide match sites (green circles). Below left are the expected fragments to be pulled down based on selective match site binding. Below right are the actual fragments pulled down (black = expected; red = unexpected; parentheses = weak). Shown top right is the polyamide with its match sequence. Shown at bottom right is the gel image for the pulldown experiment (S = supernatant; R = release).

Figure I.11. Pulldown experiments with **ATK-iix-106** and **ATK-ix-12**. Shown at left is a schematic of pUC19 with cut sites (red boxes) and polyamide match sites (green circles). Below left are the expected fragments to be pulled down based on selective match site binding. Below right are the actual fragments pulled down (black = expected; red = unexpected; parentheses = weak). Shown top right is the polyamide with its match sequence. Shown at bottom right is the gel image for the pulldown experiment (S = supernatant; R = release).

iix-129

5'-W W T G W G C W-3'

3'-W W A C W C G W-5'

Mt | NHAc | B

Expected: 322, 941 (50 single bp mismatches)

Captured: 941, **454**

pUC19 cut with MspA1 I

454, 194, 115, 309, 2335, 322, 631, 941, 518, 245, 1394, 1149

No Polyamide | 100 nM | 30 nM | 10 nM

S R S R S R S R

iix-139

5'-W W G W W G C G W-3'

3'-W W C W W C G C W-5'

NHAc | B

Expected: 194 (20 single bp mismatches)

Captured: **941**, **(518)**, 322, **(194)**

pUC19 cut with MspA1 I

454, 194, 115, 309, 2335, 322, 631, 941, 518, 245, 1394, 1149

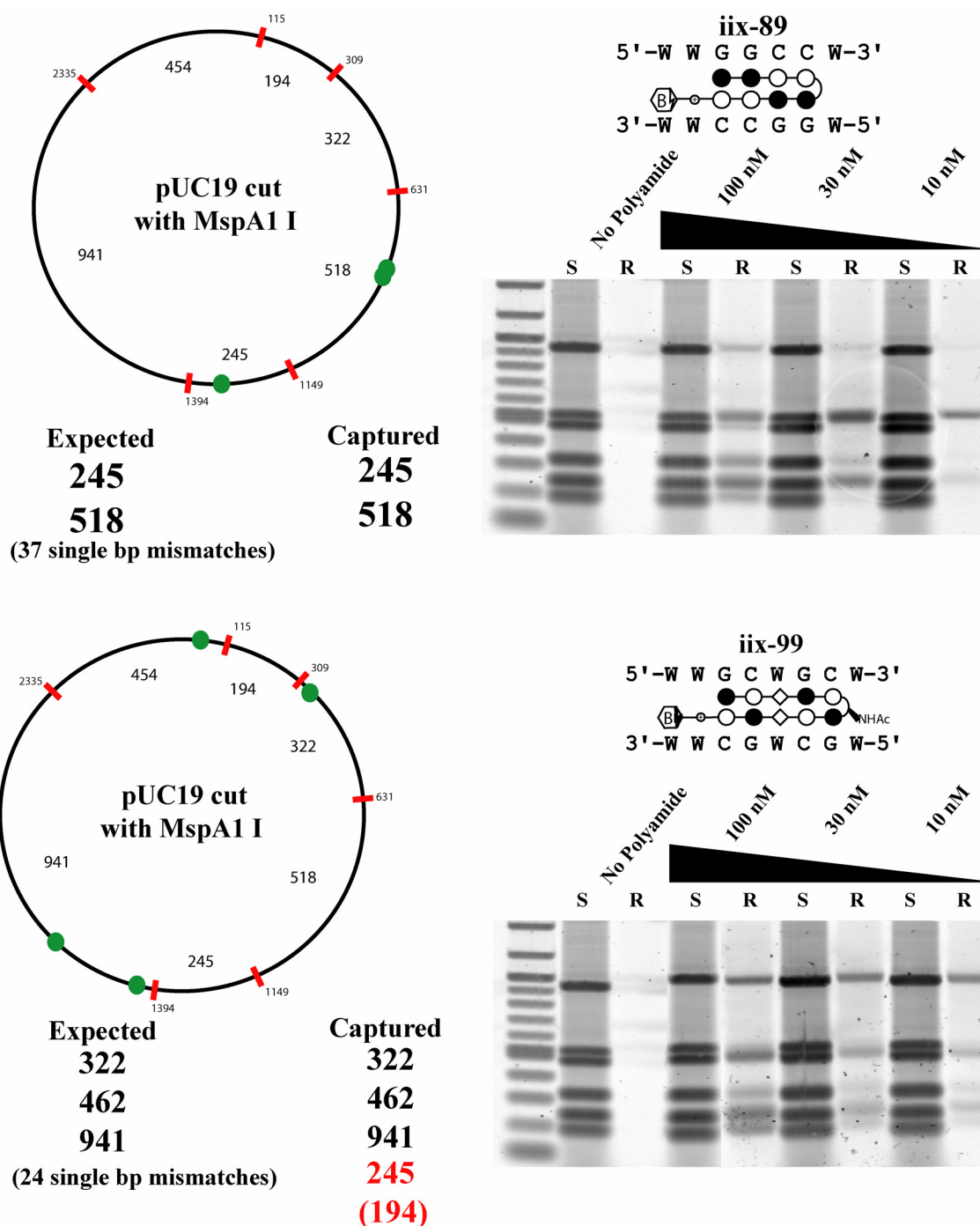No Polyamide | 100 nM | 30 nM | 10 nM

S R S R S R S R

Figure I.12. Pulldown experiments with **ATK-iix-129** and **ATK-iix-139**. Shown at left is a schematic of pUC19 with cut sites (red boxes) and polyamide match sites (green circles). Below left are the expected fragments to be pulled down based on selective match site binding. Below right are the actual fragments pulled down (black = expected; red = unexpected; parentheses = weak). Shown top right is the polyamide with its match sequence. Shown at bottom right is the gel image for the pulldown experiment (S = supernatant; R = release).

Figure I.13. Pulldown experiments with **ATK-iix-142** and **ATK-iix-143**. Shown at left is a schematic of pUC19 with cut sites (red boxes) and polyamide match sites (green circles). Below left are the expected fragments to be pulled down based on selective match site binding. Below right are the actual fragments pulled down (black = expected; red = unexpected; parentheses = weak). Shown top right is the polyamide with its match sequence. Shown at bottom right is the gel image for the pulldown experiment (S = supernatant; R = release).
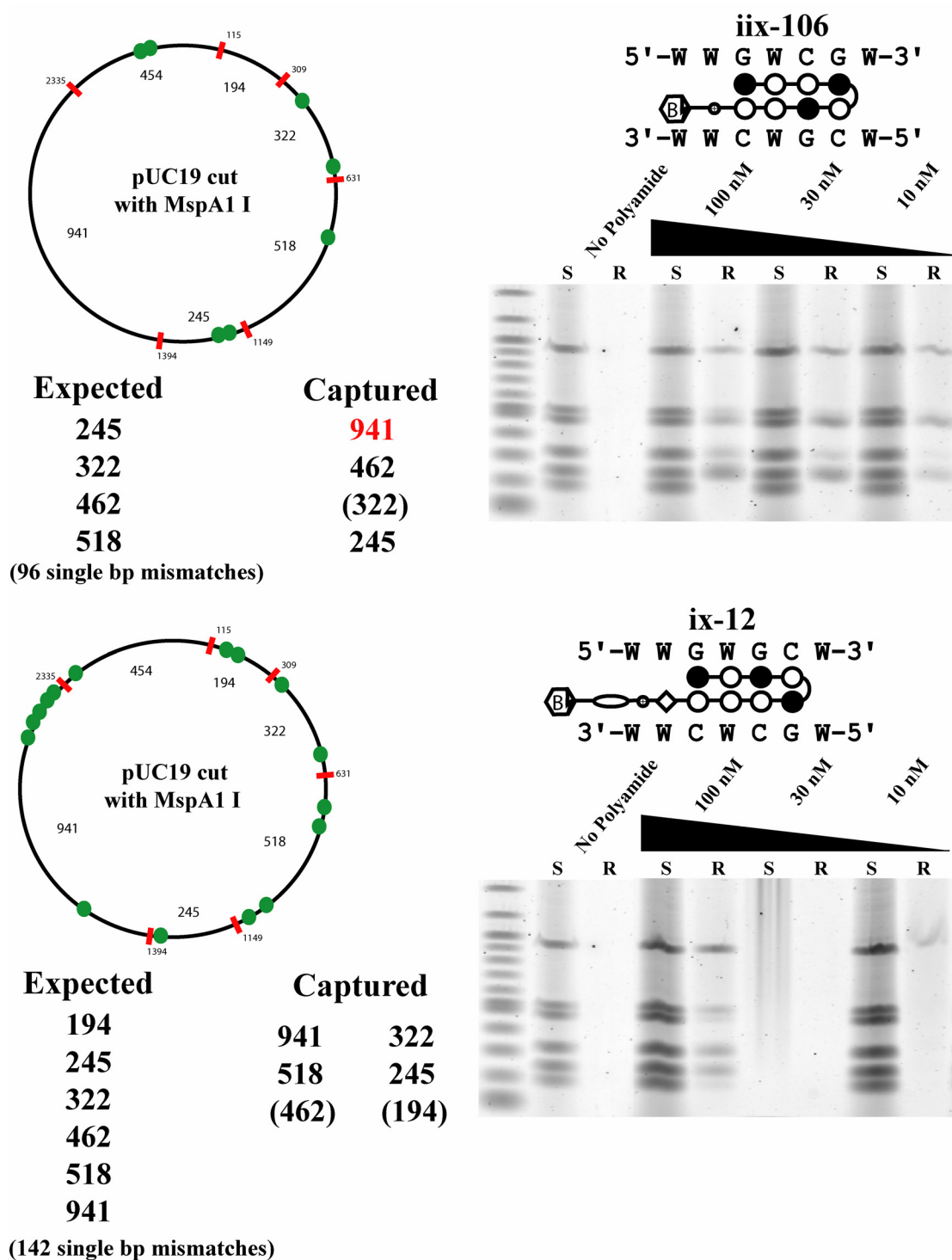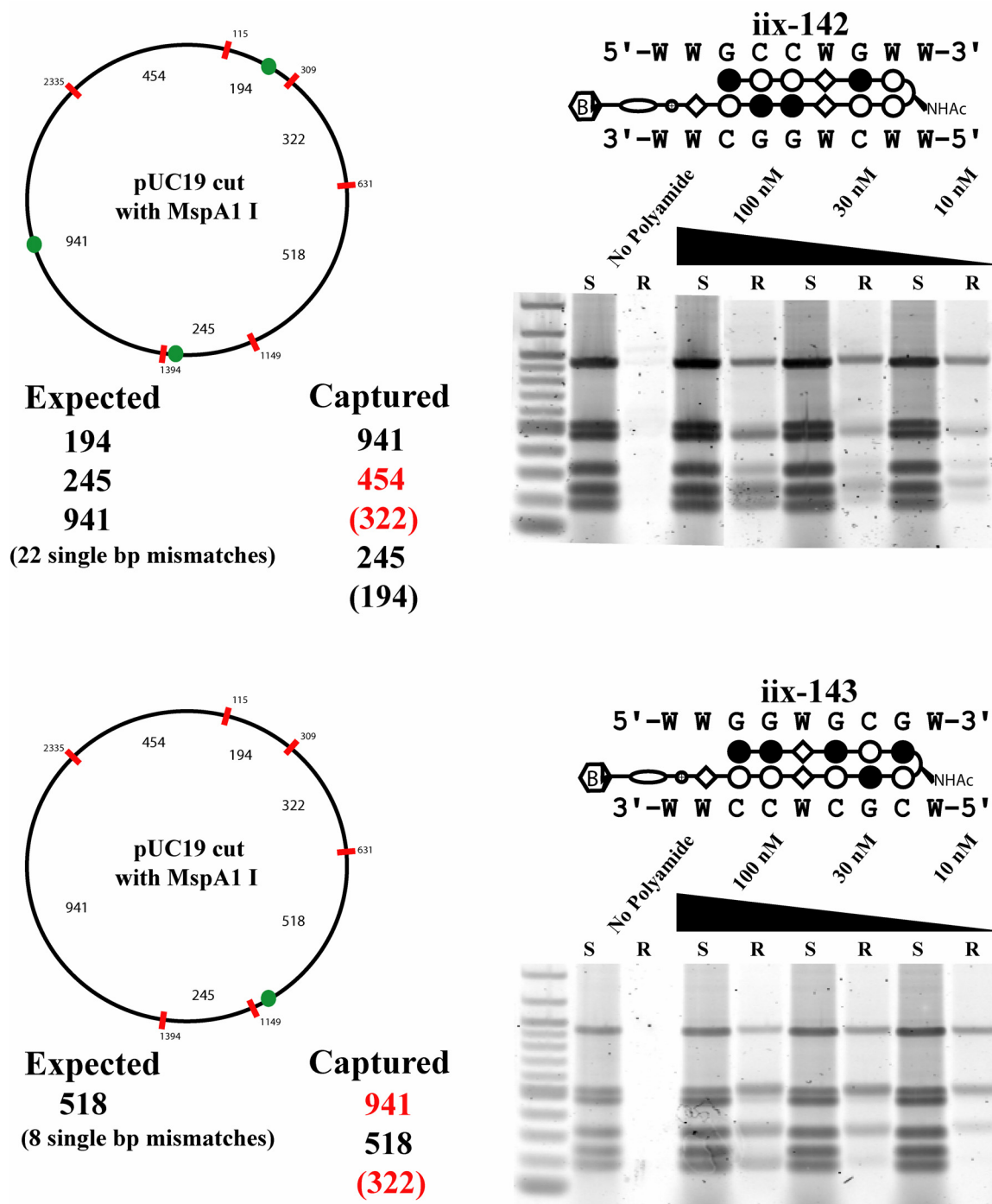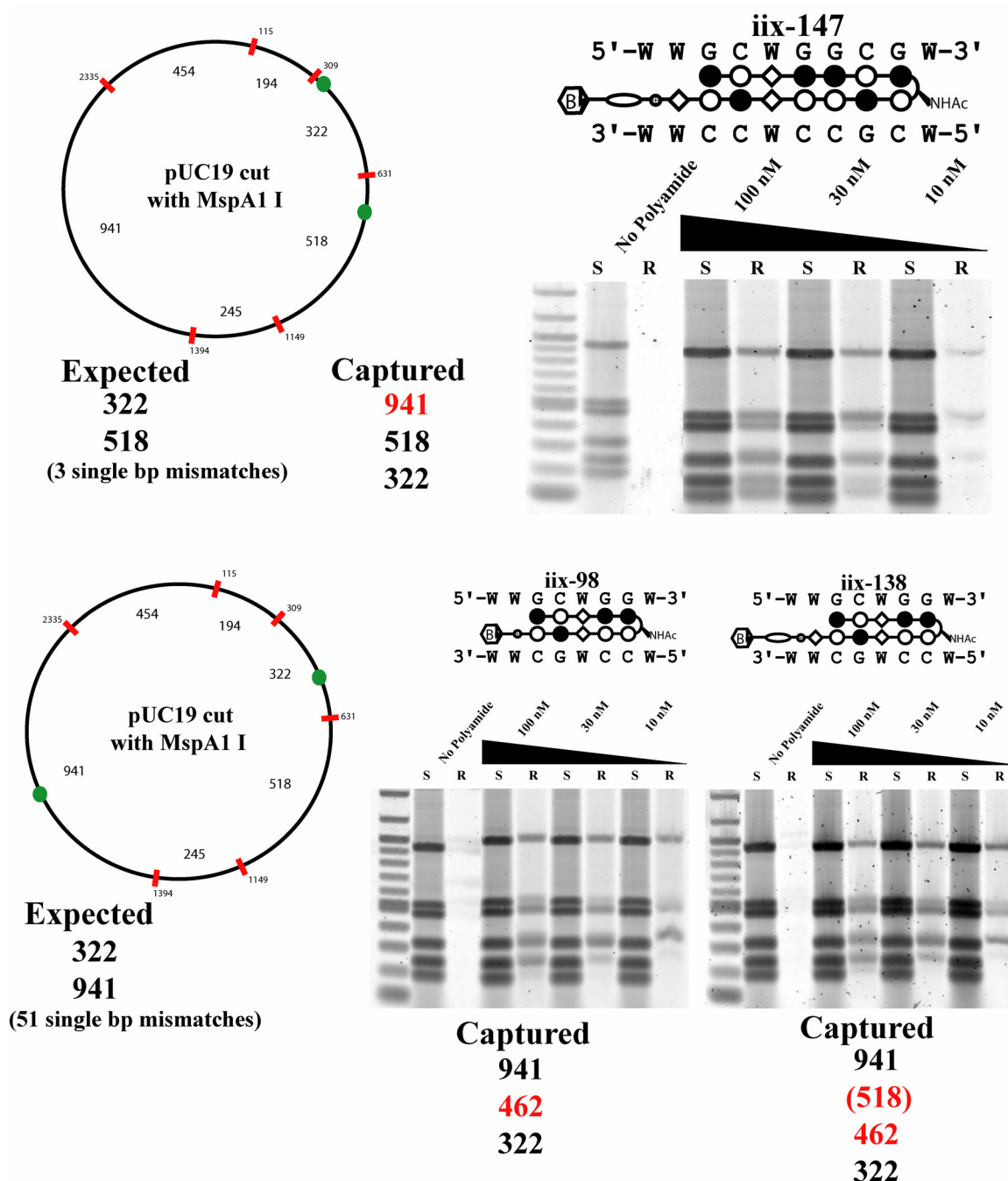
Figure I.14. Top: Pulldown experiments with **ATK-iix-147**. Bottom: Pulldown experiments with **ATK-iix-98** and **ATK-iix-138.** These two polyamides differ only by a PEG linker between the biotin and the polyamide. These data illustrate that the difference in linker length does not affect pulldown specificity. Shown at left is a schematic of pUC19 with cut sites (red boxes) and polyamide match sites (green circles). Below left are the expected fragments to be pulled down based on selective match site binding. Below right are the actual fragments pulled down (black = expected; red = unexpected; parentheses = weak). Shown top right is the polyamide with its match sequence. Shown at bottom right is the gel image for the pulldown experiment (S = supernatant; R = release).
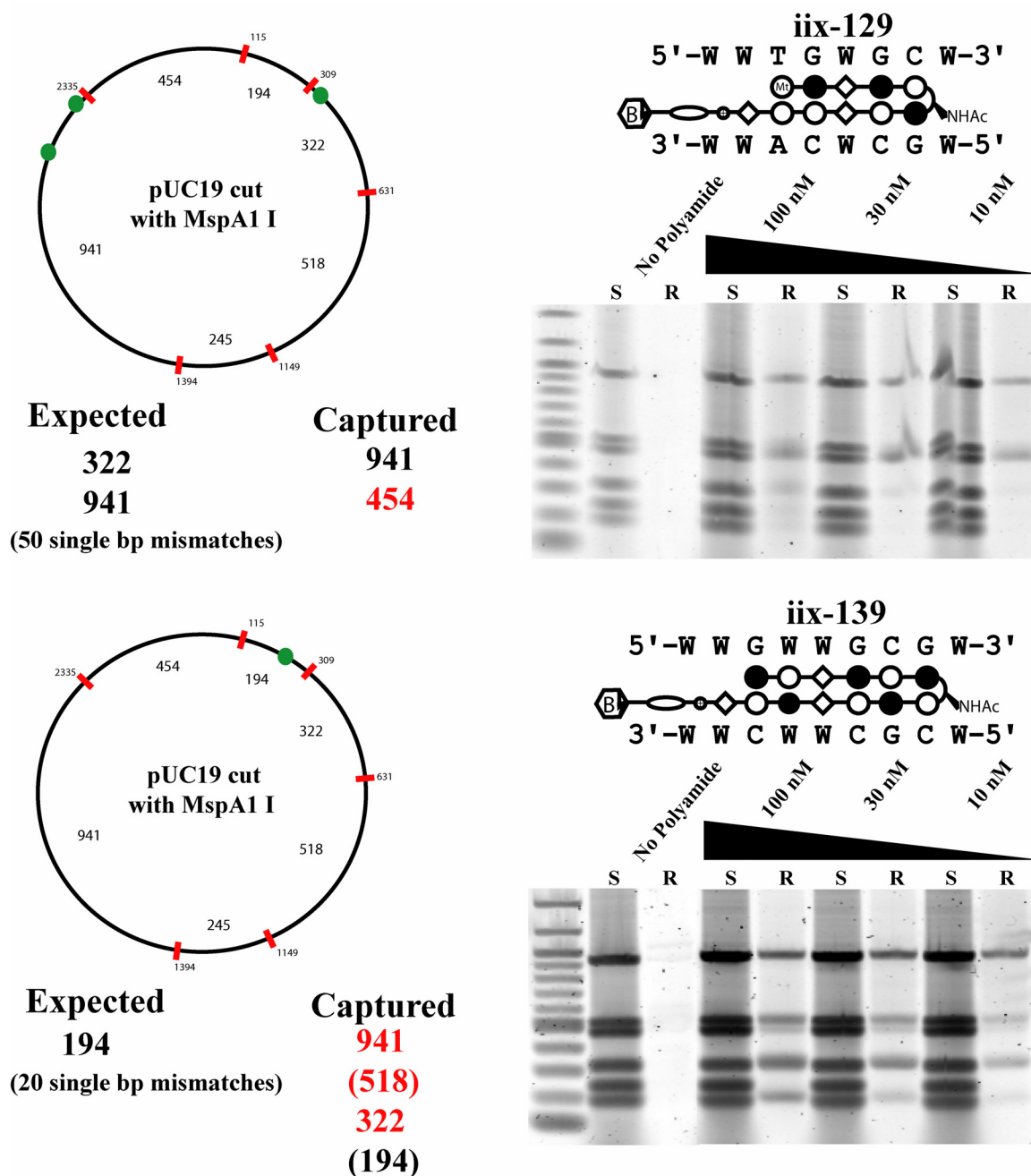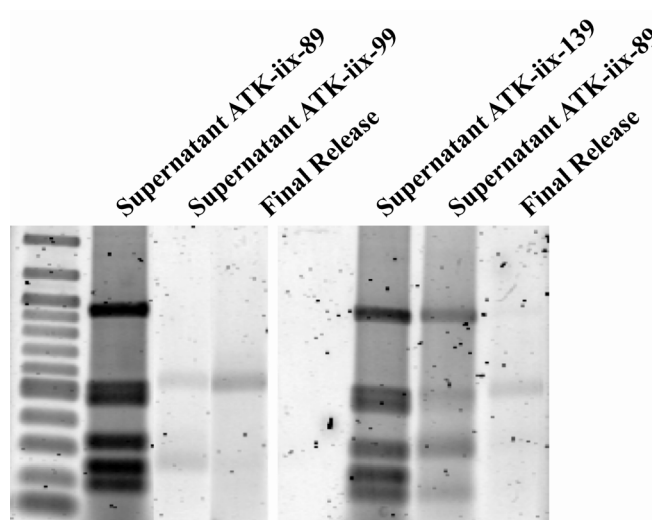
Figure I.15. Double pulldown experiments able to isolate a single DNA fragment from a complex mixture of pUC19 digestion fragments. Leftmost lane contains the DNA size ladder. Next 3 lanes- First double pulldown experiment- supernatant from 1st pulldown with ATK-iix-89; supernatant from 2nd pulldown (performed on the "release" fraction from pulldown 1) with ATK-iix-99; final release. Rightmost three lanes- Second double pulldown experiment- supernatant from 1st pulldown with ATK-iix-139; supernatant from 2nd pulldown (performed on the "release" fraction from pulldown 1); final release.

Ars1

Our collaborator (Michael Foulk; graduate student in the Susan Gerbi lab at Brown University) are interested in the autonomous replicating sequence 1 (ARS1) in the yeast genome (Figure I.16).[14] This site is one of the major origins of replications for the yeast genome, and full characterization of this 250 base pair region of DNA would be of great value to the field. The ARS1 sequence, as well as many of the factors bound to it
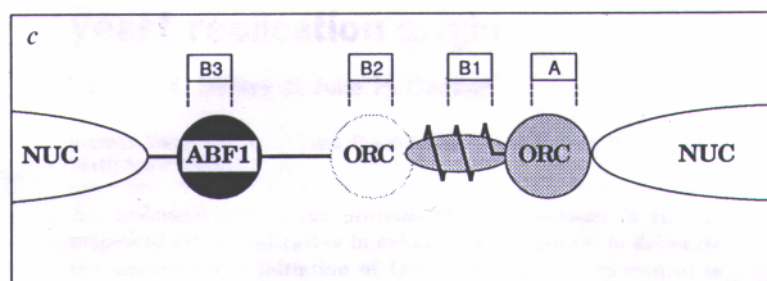

Figure I.16.[14] A plausible molecular model to explain the results of DNase I genomic footprinting at the yeast ARS1 replication start site. The origin recognition complex (ORC) is known to bind to areas A and B1. Transcription factor ABF1 binds to B3. The 250 base pair region is flanked by two tightly positioned nucleosomes.

are well-known. Using polyamide-biotin conjugates, our collaborators hope to be able to

isolate the 500 base pair restriction fragment containing the ARS1 sequence (Figure I.17)

in order to fully characterize all associated proteins. Polyamides have been designed for

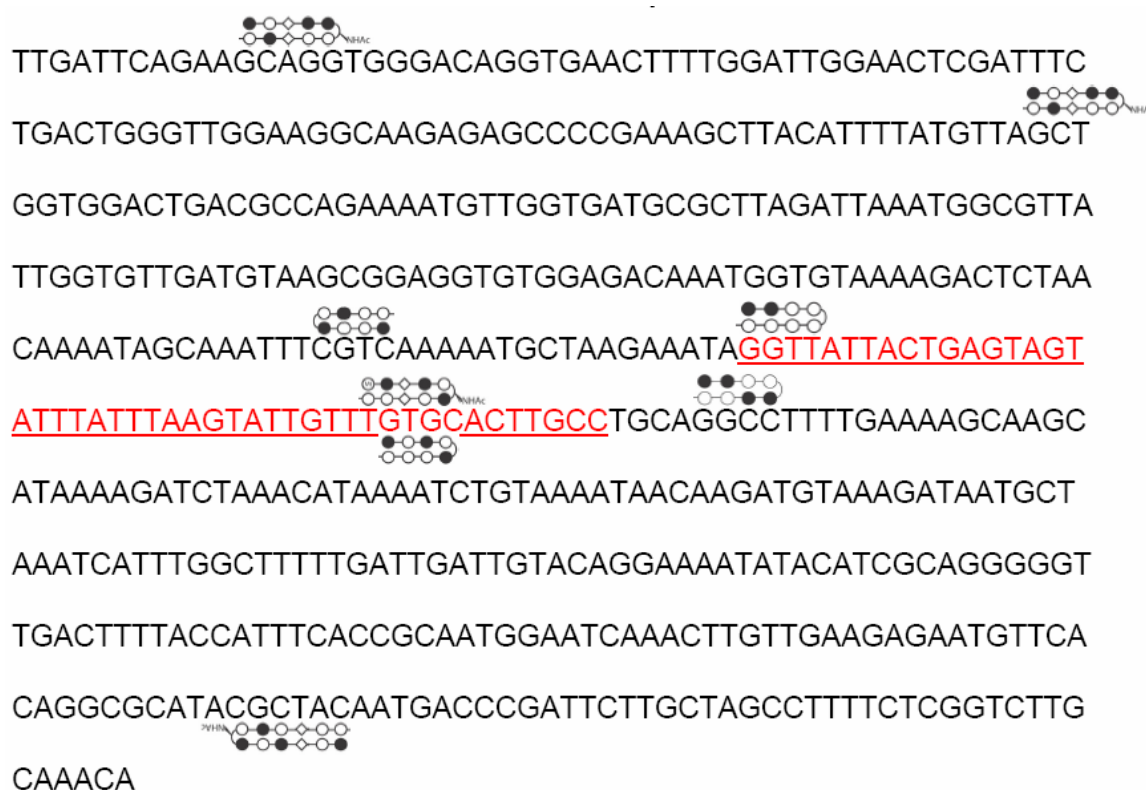this sequence (shipped to Brown University on February 11, 2004).

TTGATTCAGAAGCAGGTGGGACAGGTGAACTTTTGGATTGGAACTCGATTTC

TGACTGGGTTGGAAGGCAAGAGAGCCCCGAAAGCTTACATTTTATGTTAGCT

GGTGGACTGACGCCAGAAAATGTTGGTGATGCGCTTAGATTAAATGGCGTTA

TTGGTGTTGATGTAAGCGGAGGTGTGGAGACAAATGGTGTAAAAGACTCTAA

CAAAATAGCAAATTTCGTCAAAAATGCTAAGAAATAGGTTATTACTGAGTAGT

ATTTATTTAAGTATTGTTTGTGCACTTGCCTGCAGGCCTTTTGAAAAGCAAGC

ATAAAAGATCTAAACATAAAATCTGTAAAATAACAAGATGTAAAGATAATGCT

AAATCATTTGGCTTTTTGATTGATTGTACAGGAAAATATACATCGCAGGGGGT

TGACTTTTACCATTTCACCGCAATGGAATCAAACTTGTTGAAGAGAATGTTCA

CAGGCGCATACGCTACAATGACCCGATTCTTGCTAGCCTTTTCTCGGTCTTG

CAAACA

Figure I.17. Sequence of the 500 base pair restriction fragment containing the 250 base pair ARS1 sequence from the yeast genome. The sequence is listed 5' → 3'. The red underlined section is the region between B2 and B3 (Figure I.16) and should be most accessible to polyamides. Also shown are the polyamides designed to target this sequence (sent to the Gerbi lab, at Brown University, 2004).

Discussion.

These experiments demonstrate that fragments of DNA can be bound and isolated

using polyamide-SS-biotin conjugates. Initial pulldown experiments demonstrated that

polyamide-biotin conjugates show good specificity for their match sites. Interestingly,

AH-677-2 exhibited far less specificity than **ATK-ix-36**. Perhaps this difference in

specificity is a reflection of the fact that the imidazole ring is capable of making a

favorable hydrogen bond with the exocyclic amine of guanine, and **ATK-ix-36** contains more imidazole rings. Because these experiments are done in cell-free conditions, nuclear uptake is not an issue. Thus, larger polyamides capable of targeting long sequences of DNA can be used. In subsequent experiments, an effort was made to design polyamides with high imidazole content in order to maximize specificity. As shown by the pUC19 pulldown experiments, many of the polyamides show significant off-target pulldown. This may be due to the fact that each of the six fragments contains multiple single base pair mismatch binding sites for each of the polyamides tested. Thus, the polyamides may bind and pull down fragments with these multiple mismatches, even when no match site was present. Another issue in moving forward with DNA pulldown on a genomic scale is that as longer pieces of DNA are used, pulldown and release yields drop to under 20%. After two successive pulldowns, isolated yield is only 2%. Thus, in order for this to be feasible in a genomic setting, yields will need to be improved. Perhaps the low yields with the larger fragments of DNA were caused by size exclusion; that is, the larger DNA fragments were not tolerated by the solid support in high concentrations. The use of more magnetic beads with lower streptavidin loading may alleviate this problem. Alternatively, these experiments can possibly take advantage of multivalency. That is, polyamides can be targeted to repeat regions in the DNA. If multiple polyamide-biotin conjugates are present on a single fragment, perhaps that complex's affinity for the streptavidin-coated beads will increase, leading to improved yields. Overall, these experiments represent significant exploratory efforts towards the use of DNA-binding polyamides as tools for molecular biology.

Materials and Methods.

The base polyamides used in this study were synthesized on solid support using previously published methods and reagents.[15, 16] Biotin reagents were purchased from Pierce. Streptavidin-coated magnetic beads are from Dynal. Fluorescent oligonucleotides were purchased from IDT DNA. pUC19 was purchased from Sigma. The SYBR Gold stain was purchased from Molecular Probes.

ATK-iix-36. Base polyamide was cleaved from resin using N-methylamino dipropylamine. This was then coupled to the biotin using PFB-PEO-Biotin. UV (H_2O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1625.8. Found 1625.3.

ATK-iix-26. Synthesized according to literature procedures.[13] UV (H_2O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1946.8. Found 1946.8.

ATK-iix-47. Synthesized from base polyamide using NHS-iminobiotin (Pierce). UV (H_2O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1724.6. Found 1724.9.

ATK-iix-73. Synthesized from base polyamide using NHS-desthiobiotin (Pierce). UV (H_2O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1694.7. Found 1694.9.

ATK-iix-89. Synthesized from base polyamide using SulfoNHS-SS-Biotin (Pierce). UV (H_2O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1586.4. Found 1586.5.

ATK-iix-94. Synthesized from base polyamide using SulfoNHS-SS-Biotin (Pierce). UV (H$_2$O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1584.5. Found 1584.8.

ATK-iix-98. Synthesized from base polyamide using SulfoNHS-SS-Biotin (Pierce). UV (H$_2$O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1786.4. Found 1786.1.

ATK-iix-99. Synthesized from base polyamide using SulfoNHS-SS-Biotin (Pierce). UV (H$_2$O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1785.2. Found 1785.2.

ATK-iix-106. Synthesized from base polyamide using SulfoNHS-SS-Biotin (Pierce). UV (H$_2$O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1713.9. Found 1713.5.

ATK-iix-33 (Mono Boc-protected 1,14-(5,8,11-trioxa)undecane diamine). 1,14-(5,8,11-trioxa)undecane diamine (10 g, 58.1 mmol) was cooled to 0 ºC. Boc anhydride (1.58 g, 7.26 mmol) in 20 mL DCM was added drop wise over 30 minutes. The reaction was warmed to rt, and stirred for 1 h. 80 mL of ¼ saturated NaHCO$_3$ was then added, and the product extracted into ethyl acetate (3 x 50 mL). The combined organics were washed with saturated NaHCO$_3$ (2 x 50 mL) and brine (2 x 50 mL), dried over anhydrous magnesium sulfate, and concentrated by rotary evaporation. Product was isolated as a clear oil (2.19 g, 100%). ESI Mass [M + H] calcd. 320.4. Found 320.4.

ATK-iix-34 (BocNH-(5,8,11-trioxa)undecane succinic acid). ATK-iix-33 (2.1 g, 7.3 mmol) was dissolved in 20 mL DCM. DIEA (1.2 g, 9.6 mmol) was added and the reaction stirred for 5 minutes. Succinic anhydride (861 mg, 8.3 mmol) was then added and the reaction was stirred at room temperature overnight. The reaction was rotovapped to dryness, and the product purified by flash chromatography (10% MeOH:CHCl$_3$, rf = 0.7). Product was isolated as a white solid (1.8 g, 66%). ESI Mass [M - H]+ calcd. 419.5. Found 419.8.

ATK-iix-129. Synthesized from base polyamide, first coupling **ATK-iix-34** (1.5 equiv), PyBOP (5 equiv.), DIEA:DMF (4:1, 300 μL/μmol), followed by deprotection with 50% TFA:DCM, and final coupling with SulfoNHS-SS-Biotin (Pierce). UV (H$_2$O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 2193.8. Found 2193.7.

ATK-iix-138. Synthesized from base polyamide, first coupling **ATK-iix-34** (1.5 equiv), PyBOP (5 equiv.), DIEA:DMF (4:1, 300 μL/μmol), followed by deprotection with 50% TFA:DCM, and final coupling with SulfoNHS-SS-Biotin (Pierce). UV (H$_2$O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 2158.8. Found 2158.4.

ATK-iix-139. Synthesized from base polyamide, first coupling **ATK-iix-34** (1.5 equiv), PyBOP (5 equiv.), DIEA:DMF (4:1, 300 μL/μmol), followed by deprotection with 50% TFA:DCM, and final coupling with SulfoNHS-SS-Biotin (Pierce). UV (H$_2$O) λ_{max} 310 nm (85900). MALDI-TOF-MS calcd. (M + H): 2402.8. Found 2402.5.

ATK-iix-142. Synthesized from base polyamide, first coupling **ATK-iix-34** (1.5 equiv),
PyBOP (5 equiv.), DIEA:DMF (4:1, 300 μL/μmol), followed by deprotection with 50%
TFA:DCM, and final coupling with SulfoNHS-SS-Biotin (Pierce). UV (H_2O) λ_{max} 310
nm (85900). MALDI-TOF-MS calcd. (M + H): 2402.8. Found 2402.4.

ATK-iix-143. Synthesized from base polyamide, first coupling **ATK-iix-34** (1.5 equiv),
PyBOP (5 equiv.), DIEA:DMF (4:1, 300 μL/μmol), followed by deprotection with 50%
TFA:DCM, and final coupling with SulfoNHS-SS-Biotin (Pierce). UV (H_2O) λ_{max} 310
nm (85900). MALDI-TOF-MS calcd. (M + H): 2403.6. Found 2403.9.

ATK-iix-147. Synthesized from base polyamide, first coupling **ATK-iix-34** (1.5 equiv),
PyBOP (5 equiv.), DIEA:DMF (4:1, 300 μL/μmol), followed by deprotection with 50%
TFA:DCM, and final coupling with SulfoNHS-SS-Biotin (Pierce). UV (H_2O) λ_{max} 310
nm (1030800). MALDI-TOF-MS calcd. (M + H): 2648.5. Found 2648.7.

ATK-ix-8. Synthesized from base polyamide by coupling with SulfoNHS-SS-Biotin
(Pierce). UV (H_2O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1388.5.
Found 1388.2.

Sample Pulldown and Release Protocol.

Prepare:

5 x TKMC/T20 Buffer:
    50 mM Tris-HCl
    50 mM KCl
    50 mM $MgCl_2$
    50 mM $CaCl_2$
    pH = 7.0
    0.2 uM Filter
    Add Tween20 to the completed buffer to 0.5%

10x Solution of Polyamides in Water

50 mM DTT / 0.1% Tween 20 in Water

PROCEDURE:

1) Reactions are performed in 40 uL reaction volumes if doing one pulldown, 200 uL

reaction volumes if doing two pulldowns.

2) Set up the Polyamide/DNA equilibration (15 nM [DNA], 10-100 nM [PA])

    4 uL 10x PA solution
    8 uL 5x TKMC/T20 solution
    x uL DNA solution in water (to 15 nM)
    28-x uL H_2O
    40 uL total volume

3) Vortex to mix (1–2 sec), QuickSpin on benchtop centrifuge and place tubes in a still,

dark place to equilibrate for 3–14 hours.

4) After equilibration, set up pulldown tubes, these are new tubes with dry streptavidin-

coated beads in them. Use 100 μg beads for every 50 μL of a 50 nM solution of

polyamide. If you are far below this amount, use 50 μg of beads (that is, never use less

than 50 μg beads). To prepare the beads:

    - Take the appropriate amount of beads out of the stock solution for the total
    number of reactions.

- Dilute to 1 mL with 1x TKMC/T20.
-Using a magnet, trap the beads and remove the supernatant.
-Fill the tube with 1 mL 1x TKMC/T20.
-Remove supernatant again.
-Repeat the washing steps once more, then dilute to 1 μg/μL with 1x TKMC/T20.
-Transfer the appropriate amount of bead solution to each pulldown reaction tube (make sure that the beads are well suspended in the liquid (i.e., do not centrifuge before aliquoting).

5) Remove supernatant and add the appropriate DNA/PA solution to the washed and dried beads.

*** Be careful not to leave the beads out of solution for too long as they will not work after this.***

6) Briefly vortex the pulldown tubes to mix, centrifuge, and then set on a shaker for 20–60 min to equilibrate the biotin/streptavidin.

*** It is important in this step for the beads to remain suspended in the solution. A VWR Thermomixer shaking heat block set to 1400 rpm will accomplish this. An interval that shakes for 10 seconds and then rests for 1 min 50 seconds was found to be optimal. ***

7) When complete, centrifuge, let rest for 3–5 minutes, then, using the magnet, capture the beads and remove the supernatant into new tubes. ***This is the SUPERNATANT lane for gels. ***

8) Wash the dry beads with 40 μL 1x TKMC/T20, very briefly vortex, spin, and quickly remove the supernatant (this wash may be discarded).

9) Add 40 μL of the 1x DTT/T20 solution to the dry, washed beads and shake for 30 min-1 hour at 37 ºC using the same interval mix as in step 6.

10) When done, vortex, spin, capture beads with magnet, and remove the supernatant *** This is the RELEASE lane for gels.***

11) If you are only doing a single pulldown, skip to step 13; If you are doing multiple pulldowns proceed to step 12.

12) Take the RELEASE tubes and do a phenol:chloroform extraction, keeping the aqueous layer.

13) To all tubes (SUPERNATANT and RELEASE) add 2 volumes EtOH, invert to mix, and spin at 14000 rpm at 4 ºC for 30 minutes to precipitate the DNA.

14) Remove the supernatant. If doing multiple pulldowns, skip to step 17.

15) Let air dry for 15 min, then take up in 20 μL H_20. Dissolve the pellet by vortexing (5 sec), spinning, then letting the tubes sit for 15 min.

16) Add 5 uL 10x Ficol/TBE (to final conc. of 2x) and load on an agarose gel. Run gel. Visualize with SYBR Gold stain using Typhoon Phosphorimager.

For multiple pulldowns

17) To the RELEASE tubes from step 10 (carried through steps 11–14) Let the tubes air dry 15 minutes then take up the pellet in a new solution of 1 x TKMC/T20 and polyamide. Because of the low pulldown yield, I usually use the same concentration of polyamide as I did in the first pulldown reaction but I decrease the volume by ~4-fold. This solution then becomes your second equilibration reaction. It can then be carried through steps 3 to the end.

References.

1. Bernstein, B. E.; Humphrey, E. L.; Liu, C. L.; Schreiber, S. L., The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. In *Chromatin and Chromatin Remodeling Enzymes, Pt B*, 2004; Vol. 376, pp 349–360.

2. Buck, M. J.; Lieb, J. D., *Genomics*, **2004,** *83*, 349–360.

3. Ciccone, D. N.; Morshead, K. B.; Oettinger, M. A., Chromatin immunoprecipitation in the analysis of large chromatin domains across murine antigen receptor loci. In *Chromatin And Chromatin Remodeling Enzymes, Pt B*, 2004; Vol. 376, pp 334–367.

4. Das, P. M.; Ramachandran, K.; vanWert, J.; Singal, R., *Biotechniques*, **2004,** *37*, 961–969.

5. Ren, B.; Dynlacht, B. D., Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors. In *Chromatin And Chromatin Remodeling Enzymes, Pt B*, 2004; 'Vol.' 376, pp 304–312.

6. Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology*, **2003,** *13*, 284–299.

7. 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.

8. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Nature*, **1998,** *391*, 468–471.

9. Suto, R. K.; Edayathumangalam, R. S.; White, C. L.; Melander, C.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., *Journal of Molecular Biology*, **2003,** *326*, 371–380.

10. Carter, S. A.; *Ph. D. Thesis;* California Institute of Technology, Pasadena, **1998**.

11. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. Q. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J. H.; Miklos, G. L. G.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, C.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z. M.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W. M.; Gong, F. C.; Gu, Z. P.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z. X.; Ketchum, K. A.; Lai, Z. W.; Lei, Y. D.; Li, Z. Y.; Li, J. Y.; Liang, Y.; Lin, X. Y.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B. X.; Sun, J. T.; Wang, Z. Y.; Wang, A. H.; Wang, X.; Wang, J.; Wei, M. H.;

Wides, R.; Xiao, C. L.; Yan, C. H.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W. Q.; Zhang, H. Y.; Zhao, Q.; Zheng, L. S.; Zhong, F.; Zhong, W. Y.; Zhu, S. P. C.; Zhao, S. Y.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H. J.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H. Y.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A. D.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X. J.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M. Y.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. H., *Science*, **2001,** *291*, 1304–1319, and suppliments.

12. Heckel, A.; *Postdoctoral Report;* California Institute of Technology, Pasadena, **2002**.

13. Rothschild, K. J.; Sonar, S. M.; Olejnik, J. Photocleavable agents and conjugates having detectable moieties and photoreactive moieties for the detection and isolation of biomolecules. *U.S. Patent*; **1999**.

14. Bielinsky, A.-K.; Gerbi, S. A., *Science*, **1998,** *279*, 95–98.

15. Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6141–6146.

16. Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 2767–2774.

Appendix II

Controlling the Binding Orientation of Tailless Hairpin Polyamides

Abstract

Most hairpin polyamides exhibit a binding orientation preference aligning N-terminal to C-terminal along the DNA in the 5' to 3' direction. It has been postulated that this orientation preference is caused by the C-terminal alkyl tail of the polyamide molecule. With recent advances in the solid-phase synthesis of polyamides, molecules without C-terminal tails are now accessible. We find that these tailless polyamides exhibit severely reduced orientation preference when compared to their tail-containing counterparts. We find that installation of (R)-diaminobutyric acid at the turn position is able to restore preference for "forward" binding. However, installation of the (S)-diaminobutyric acid does not favor "reverse" binding.

Introduction.

Pyrrole-imidazole hairpin polyamides bind the minor groove of DNA with high affinity and sequence specificity.[1] Most hairpin polyamides synthesized to date bind with the N-terminus oriented towards the 5' end of the DNA.[2] It has been postulated that the C-terminal alkyl tail is responsible for the orientation preference of minor groove-binding polyamides.[3] Indeed, the natural products distamycin and netropsin closely resemble polyamides in structure (Figure II.1). Structural studies have found that distamycin possesses a strong orientation preference when bound in 2:1 complexes with polyA-DNA, favoring
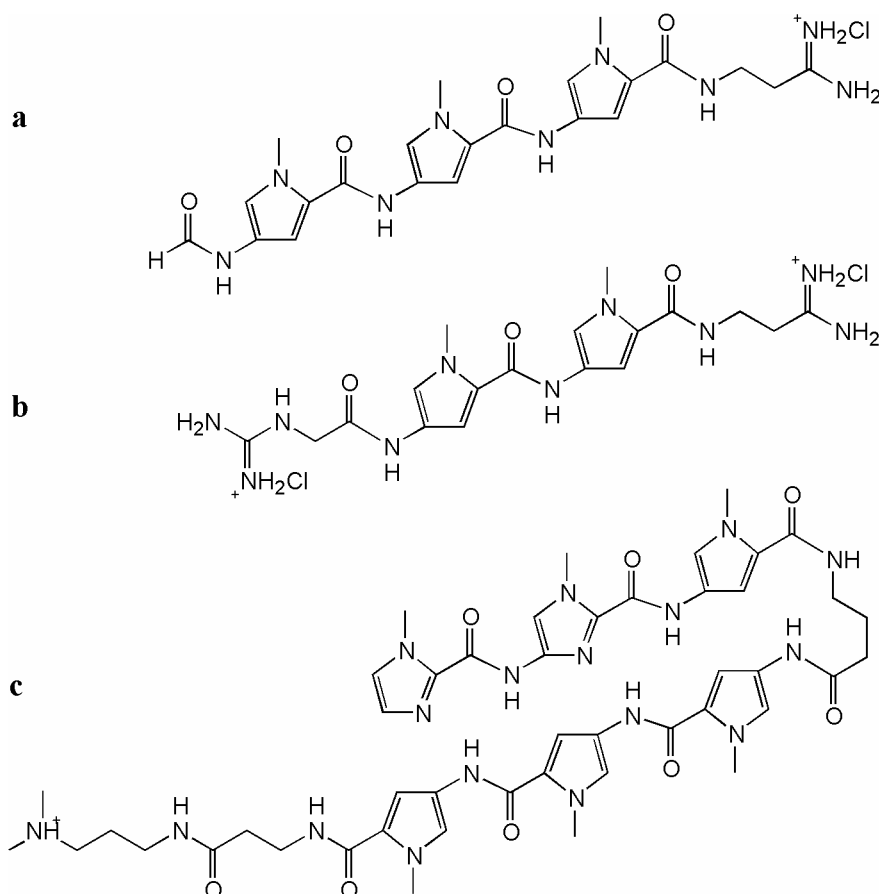

Figure II.1. Structures of minor groove-binding agents distamycin (a); netropsin (b); pyrrole-imidazole polyamide with β-Dp tail (c).

N-terminal alignment with the 5' end of the DNA by greater than 20:1.[4-6] Conversely, netropsin bound to the same sequence of DNA exhibits only a 3:2 orientation preference.[7] Indeed, the most obvious structural difference between these two molecules is that netropsin has an alkyl tail at both its C- and N-termini, while distamycin has only a C-terminal tail.

NMR structures of a 6-ring hairpin polyamide bound to the minor groove in both the "forward" (N→C aligning 5'→3') and "reverse" (N→C aligning 3'→5') orientations reveal interesting structural characteristics. All structures reveal a twist in the ligand that
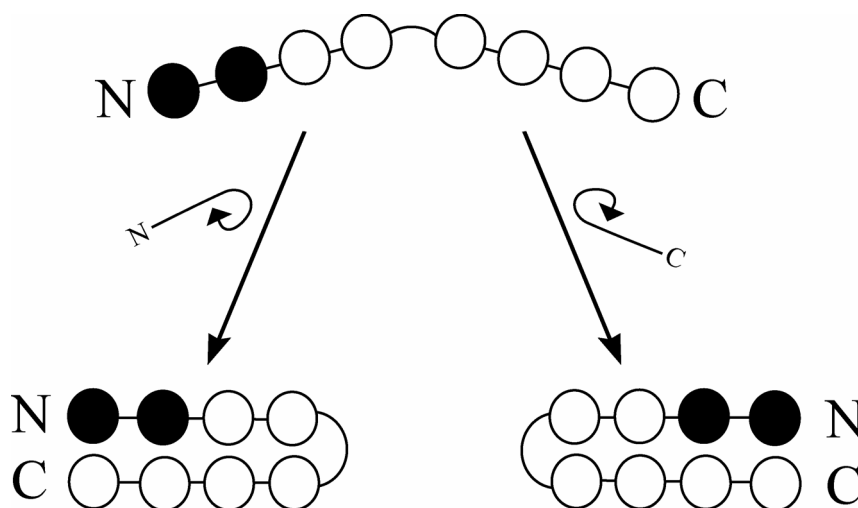

Figure II.2.[8] Model for the two folding geometries of a hairpin polyamide. Folding pathways lead to hairpin structures suitable for recognition of DNA in the forward orientation (left) and reverse orientation (right). Pyrrole and imidazole carboxamides are represented by white and black circles, respectively.

allows it to better match the curvature of the DNA. When bound in the forward orientation, the twist allows the C-terminal tail to lie along the floor of the groove. Conversely, when bound in the reverse orientation, the twist would cause the C-terminal tail to sterically clash with the wall of the minor groove, and as a result, the tail points out into solution in order to avoid this clash.[3] Thus, when bound in the reverse orientation,

polyamides lose the favorable hydrophobic interactions between the alkyl tail and the hydrophobic minor groove.

For any given hairpin polyamide core, there exist two, non-superimposable hairpin folds that are related by mirror plane symmetry (Figure II.2).[8] The binding of these two folds to DNA is energetically distinguished by the tail effects described above, thus leading to the experimentally observed orientation preference. With the recent advances in solid-phase synthesis methodology,[9] tailless polyamides are now synthetically accessible. Because the orientation preference of polyamides is postulated to be a direct consequence of the presence of the C-terminal tail, these tailless polyamides may not have any orientation preference.

Recent studies have shown that installation of a chiral amine on the γ-aminobutyric acid "turn" residue acts as an additional determinant for orientation preference. When polyamides are functionalized with the (R)-enantiomer of γ-aminobutyric acid they possess high affinity and specificity for forward orientation binding. When the (S)-enantiomer is installed, binding affinity at the forward orientation match site is reduced almost 200-fold (note, the compound still possesses a slight preference for the forward binding site due to the presence of a C-terminal β-Dp tail).[8] Computer-generated models show that the differences in binding are caused by differential placement of the amine functionality. In the forward orientation, the (S)-enantiomer causes the amine to sterically clash with the wall of the minor groove, while the (R)-enantiomer allows the amine to point freely out of the minor groove.[8]

In this study, we are interested in whether the enantiomers of the γ-aminobutyric acid turn can control the binding orientation of tailless hairpin polyamides. In the

absence of a C-terminal tail, the chiral turn should be the only determinant of binding orientation. As shown in figure II.3, the (R)-chiral turn should favor the forward orientation, but disfavor reverse orientation binding, while the (S)-chiral turn should favor reverse binding and disfavor the forward orientation.
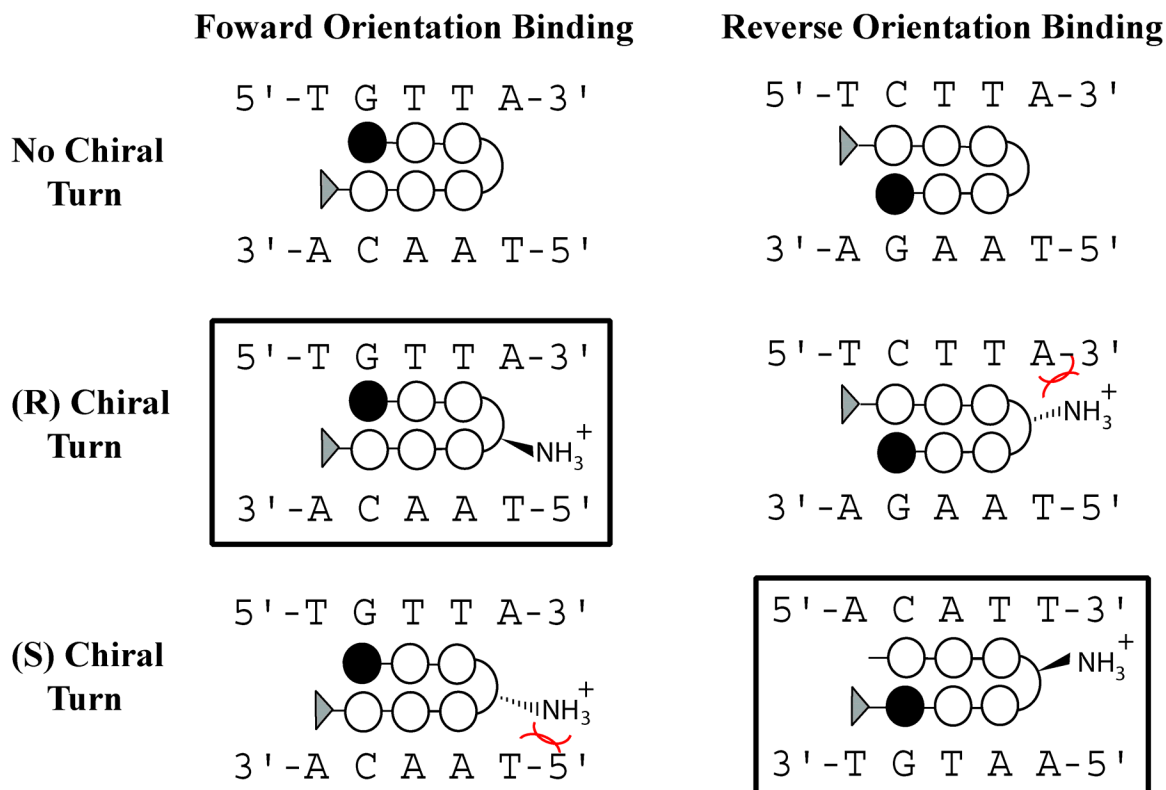

Figure II.3. Schematic representation of how the chiral turn may be used to control binding orientation of tailless hairpin polyamides. Top row, unfunctionalized tailless polyamides should not possess any orientation preference. Middle row, Functionalization with the (R)-chiral turn should favor the forward binding orientation (boxed) and lead to a steric clash with the wall of the minor groove when bound in the reverse orientation. Bottom row, Functionalization with the (S)-chiral turn should favor the reverse binding orientation (boxed) and lead to a steric clash with the wall of the minor groove when bound in the forward orientation.

Results.

Polyamides **ATK-iix-73**, **ATK-iix-74**, **ATK-iix-75**, **ATK-iix-76**, and **ATK-iix-77** were synthesized on oxime resin using standard reagents and procedures (Figure

II.4).[9] Each polyamide was designed to contain a single positive charge. Compounds **ATK-iix-74** and **ATK-iix-76** are analogs of **ATK-iix-73** and **ATK-iix-75**, respectively, in which the chiral amine has been acylated. The acyl group provides additional steric bulk in a stereocontrolled manner, and should accentuate any orienting effects of the chiral group. Compound **ATK-iix-77** is the control compound with no functionality at
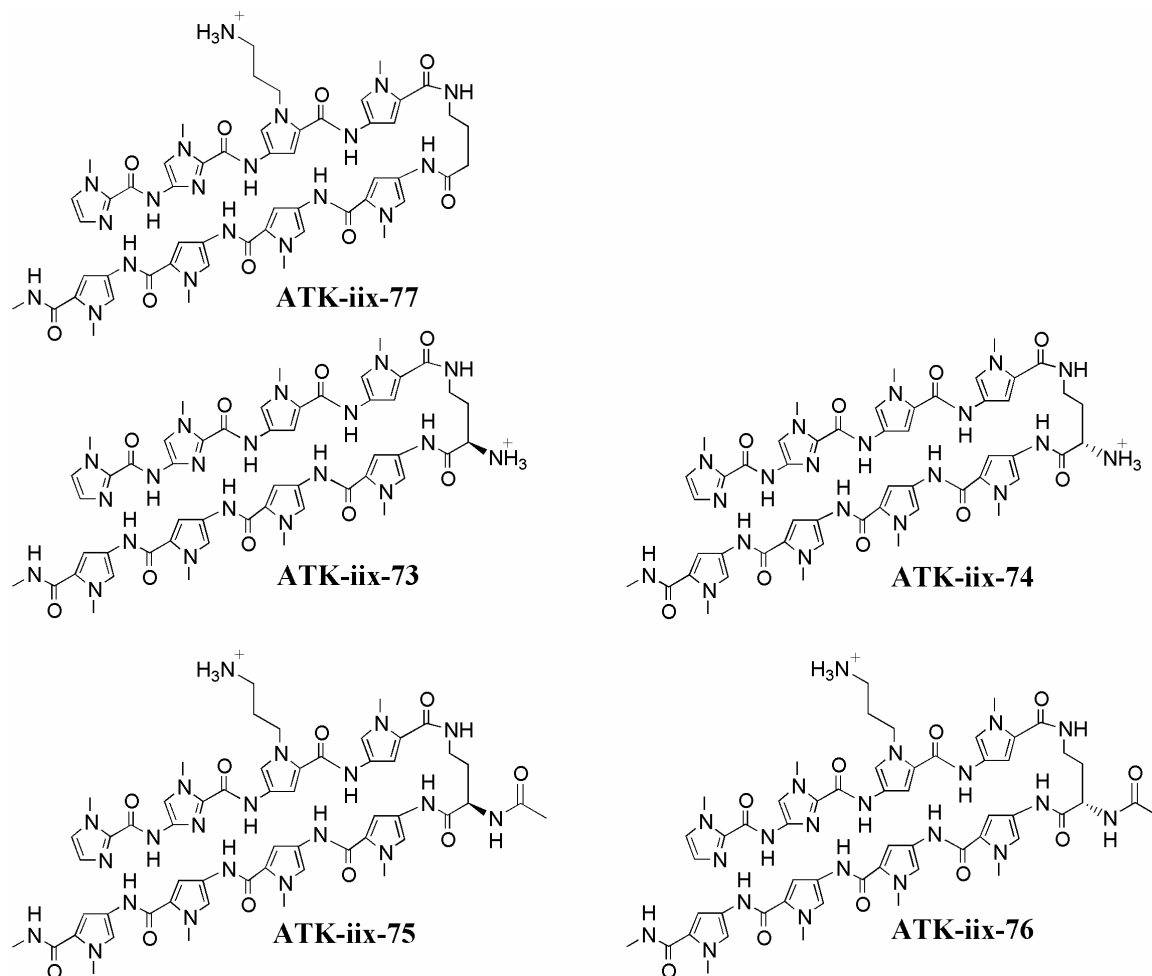

Figure II.4. Chemical structures of the hairpin polyamides synthesized for this study. Each polyamide is designed to recognize the sequence WGGWWW (where W = A or T). Each polyamide has a single positive charge.

the turn position. The β-Dp analog of **ATK-iix-77** has previously been footprinted, and binds with an association constant of $1.9 \times 10^9 \text{ M}^{-1}$ (at 5'-TGGTAT-3'), exhibiting >500-fold preference for forward orientation binding.[10]

Equilibrium association constants were determined on the ^{32}P-labeled restriction fragment from pATK4, which contains the inserted sequence shown in figure II.5. The insert contains match and single base pair mismatch sites for both forward and reverse binding orientations. DNase I footprinting results are shown below (Figures II.6–8). Control compound **ATK-iix-77** binds to the forward match site with an affinity of 3.1 x 10^9 M^{-1}. This compound also binds the reverse match site with an affinity of 4.1 x 10^8 M^{-1}, thus favoring forward binding by 7.5-fold.
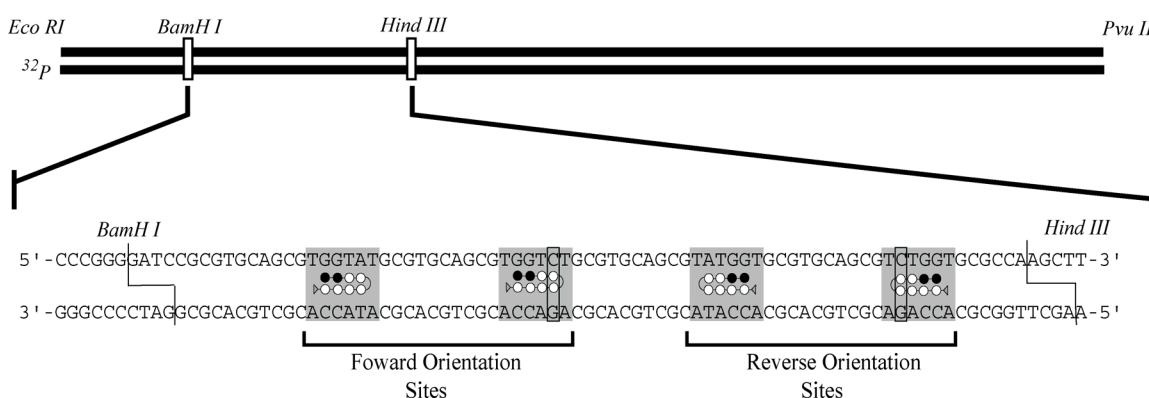

Figure II.5. Diagram of the restriction fragment insert from pATK4 used for DNase I footprinting assays. The insert contains the forward orientation match site (left), a single base pair mismatch for the forward orientation site (center-left), the reverse orientation match site (center-right), and a single base pair mismatch for the reverse orientation site (right).

Compounds possessing the (R) stereochemistry both bind the forward match site with high affinity (5.1 x 10^9 M^{-1} and 6.3 x 10^9 M^{-1}, for **ATK-iix-73** and **ATK-iix-74**, respectively). These polyamides show reduced binding at the reverse match site, with association constants of 1.4 x 10^8 M^{-1} and 5.9 x 10^7 M^{-1}, for **ATK-iix-73** and **ATK-iix-74**, respectively.

The (S)-amine compound **ATK-iix-75** binds to the forward match site with 1.2 x 10^7 M^{-1} affinity and the reverse match site with an association constant < 3 x 10^6 M^{-1}. The (S)-acylated amine compound **ATK-iix-76** does not show binding at either site up to 300 nM concentrations. The DNase I footprinting results are summarized in table II.1.
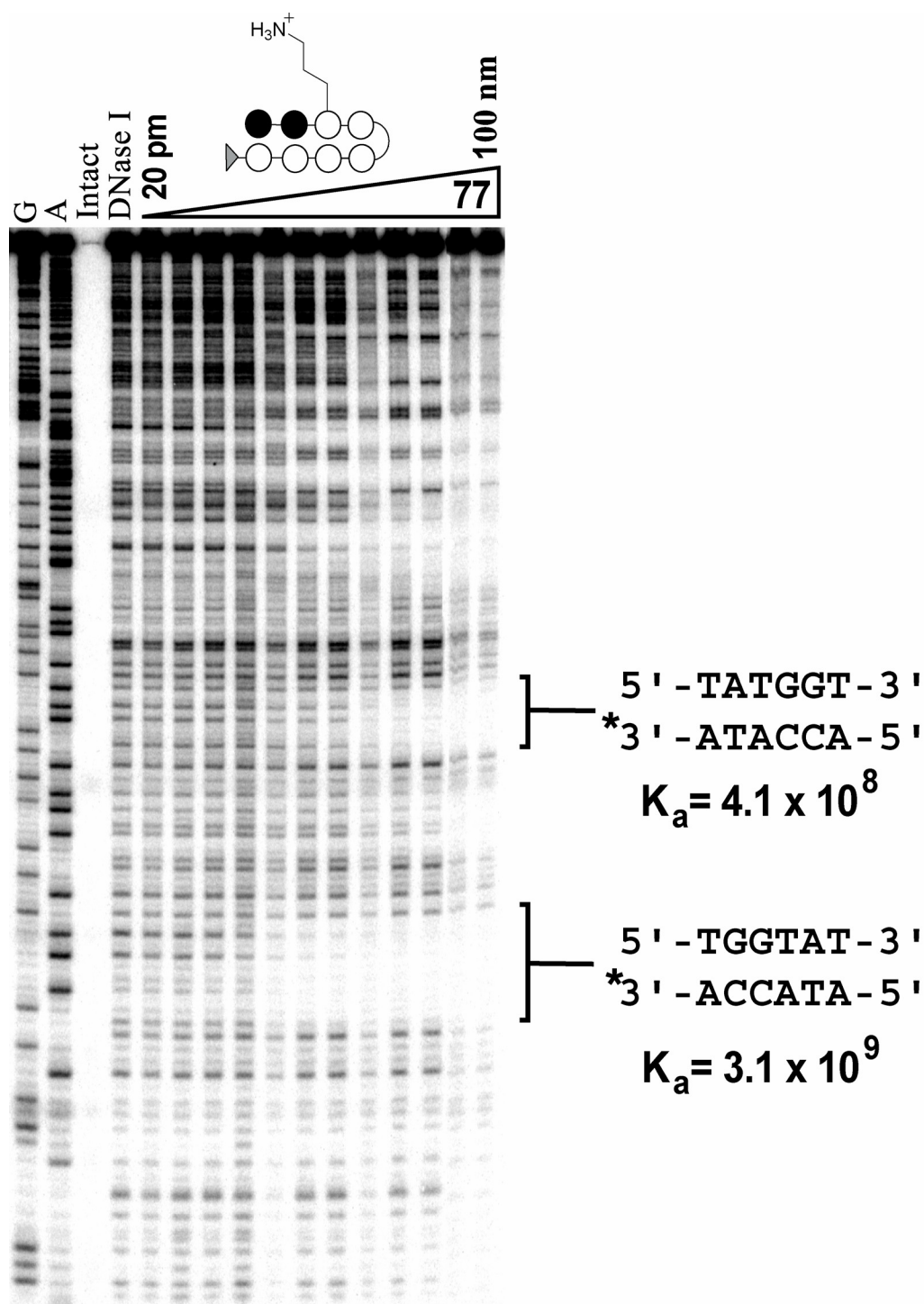
Figure II.6. DNase I footprinting assay on pATK4. Lanes from left to right are G sequencing reaction, A sequencing reaction, intact DNA, lanes 4–15 contain 0, 20pM, 50 pM, 100 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM concentrations of **ATK-iix-77**. Equilibrium association constants shown at right.
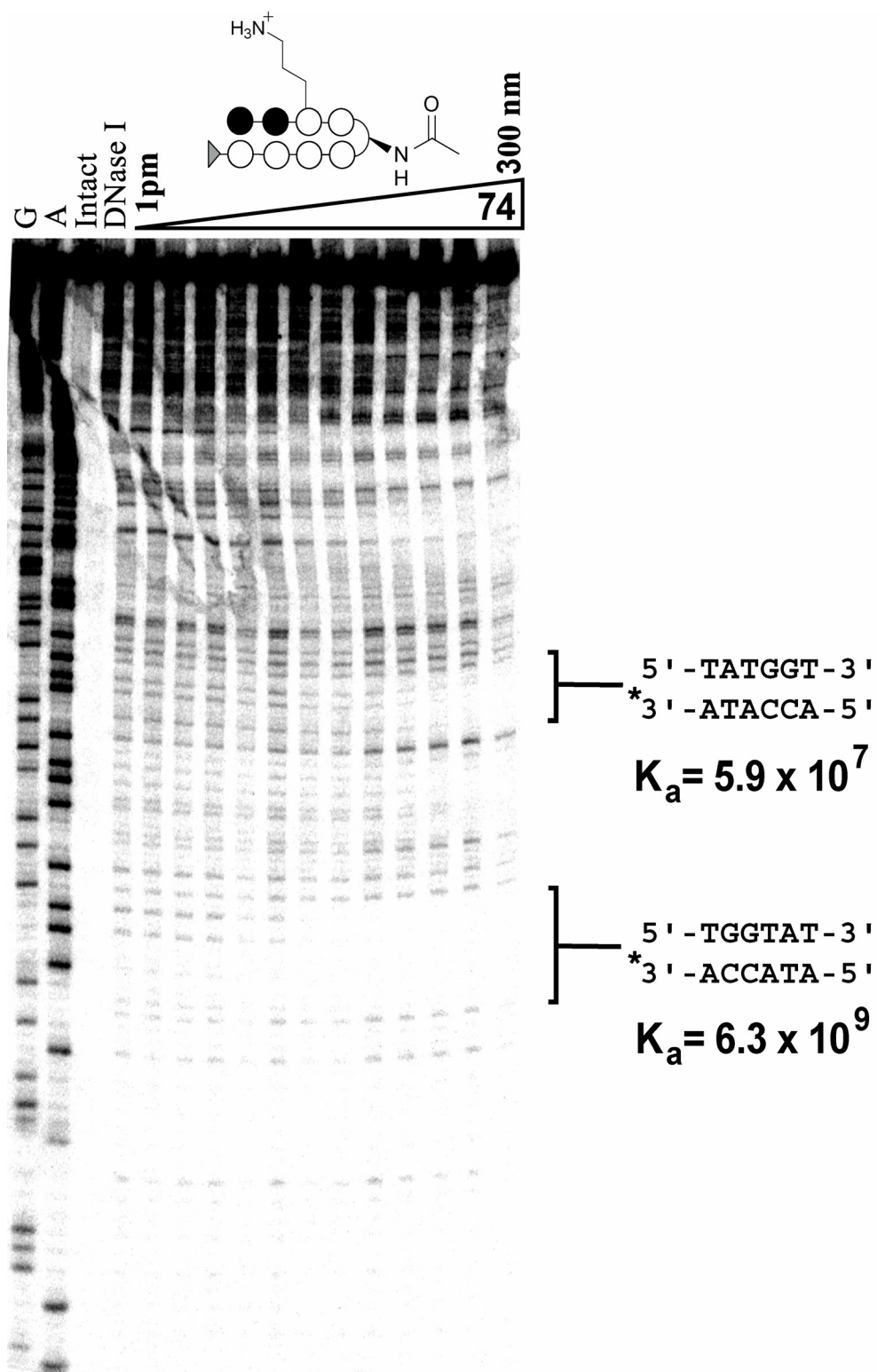
Figure II.7. DNase I footprinting assay on pATK4. Lanes from left to right are G sequencing reaction, A sequencing reaction, intact DNA, lanes 4–16 contain 0, 20pM, 50 pM, 100 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 300 nM concentrations of **ATK-iix-74**. Equilibrium association constants shown at right.

Figure II.8. DNase I footprinting assay on pATK4. Lanes from right to left are A sequencing reaction, G sequencing reaction, intact DNA, lanes 4–16 contain 0, 20pM, 50 pM, 100 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 300 nM concentrations of **ATK-iix-75**. Equilibrium association constants shown at right.
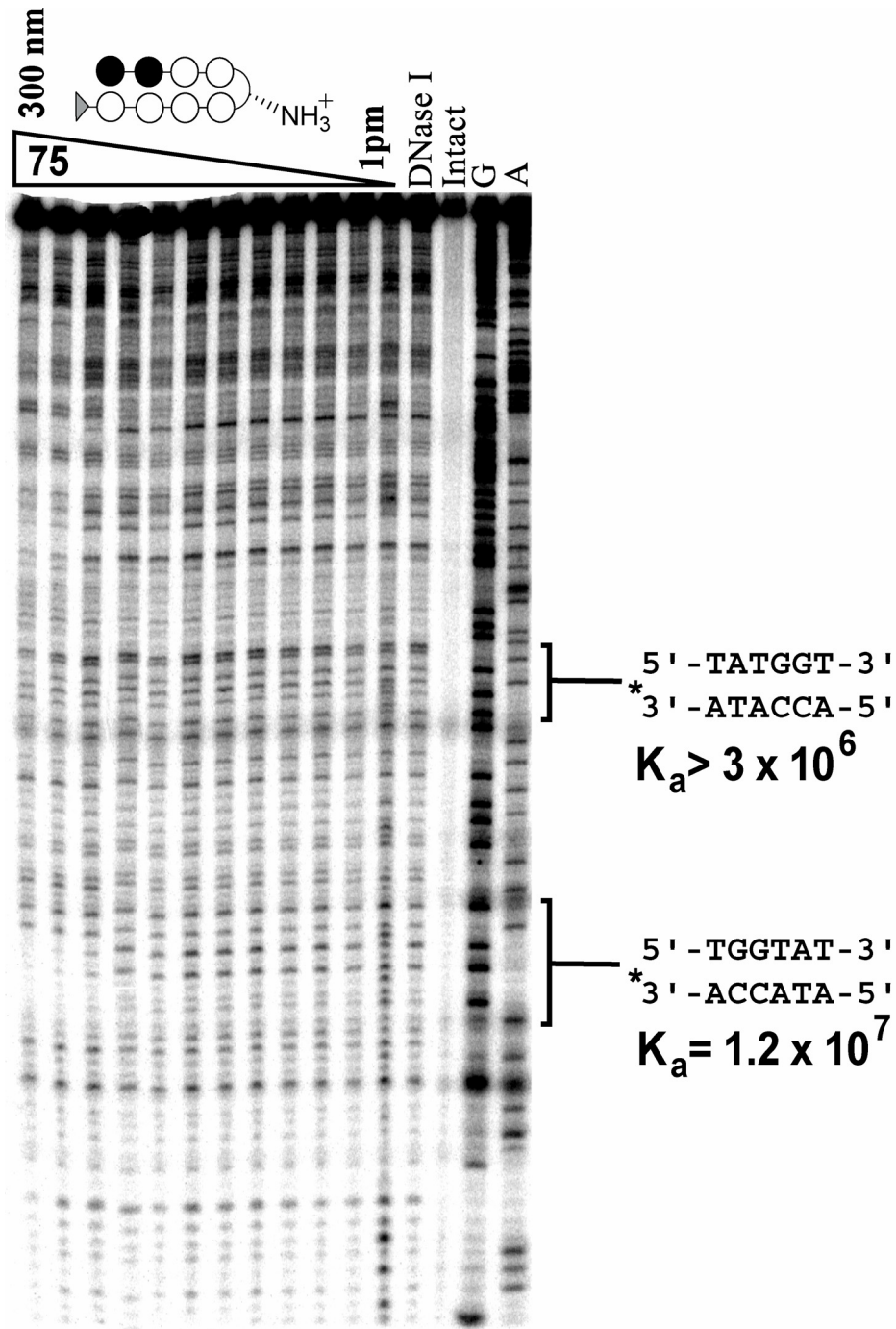
Table II.1. Affinity association constants for tailless polyamides.[a]

Polyamide	Forward Match Site 5′-TGGTTA-3′	Reverse Match Site 5′-ATTGGT-3′	Forward Orientation Preference (5'→ 3')[b]
ATK-iix-77	3.1×10^9	4.1×10^8	7.5
ATK-iix-73	5.1×10^9	1.4×10^8	36
ATK-iix-74	6.3×10^9	5.9×10^7	107
ATK-iix-75	1.2×10^7	$< 3 \times 10^6$	>4
ATK-iix-76	$< 3 \times 10^6$	$< 3 \times 10^6$	--

a. Each association constant is the average of three quantitative footprint titrations. Values are reported in units of M^{-1}. Standard deviations are no more than 8% of each reported value. b. Orientation preference calculated as the ratio of the binding affinity at the forward match site versus the reverse match site.

Conclusion.

With the advent of solid-phase synthesis methodologies allowing for the synthesis of hairpin polyamides with truncated C-terminal tails, we were interested in whether these tailless polyamides exhibit any binding orientation preference. Also, we were interested in whether the orientation preference could be controlled or reinforced by the introduction of a chiral amine at the turn position.

Tailless polyamide **ATK-iix-77** shows good binding affinities to both the forward and reverse match sites. This tailless polyamide exhibits only a 7.5-fold preference for the forward site (over the reverse site). This is in stark contrast to the β-Dp analog of this compound, which exhibits a 500-fold orientation preference. Specificities over single base pair mismatches in both orientations were >100-fold.

The (R) stereochemistry at the turn should favor the forward binding orientation. Indeed, when the (R) amine is installed (**ATK-iix-73**), the compound shows increased affinity for the forward match site and reduced affinity at the reverse match site (with

respect to unfunctionalized control **ATK-iix-77**). This results in a 36-fold preference for the forward orientation. When the bulk of the chiral group is increased to the acylated amine (**ATK-iix-75**), the affinity at the reverse binding site is further decreased. This compound exhibits a 106-fold preference for the forward binding site. Thus, the (R) chiral turn acts as a determinant for "forward" orientation preference. Indeed, it is able to restore "forward" orientation preference to tailless polyamides.

Neither of the compounds possessing the (S)-stereochemistry is able to bind any of the sites with reasonable affinities. Indeed, only the (S)-amine compound **ATK-iix-74** showed any binding at all. While this enantiomer was hypothesized to favor the reverse binding orientation, the compound favors the forward site by 4-fold.

We have shown that tailless polyamides exhibit severely decreased orientation preference with respect to the C-terminal β-Dp analogs. Clearly, the tail functionality does account for some of the orientation preference exhibited by hairpin polyamides. Installation of a chiral turn with (R)-stereochemistry is able to restore the orientation preference for the forward binding orientation for tailless hairpin polyamides. Conversely, use of a chiral turn with (S)-stereochemistry is unable to force reverse binding.

Thus, while complete control of binding orientation of tailless polyamides was not achieved, it was found that researchers should be careful to install an (R)-chiral amine or acylated amine on tailless polyamides in order to obtain molecules that are specific for the forward orientation match site over the reverse orientation match site.

Materials and Methods.

Polyamides were synthesized on solid support as previously described.[9] Propylamino

pyrrole was installed as previously described.[11]

ATK-iix-73. Synthesized on solid support and cleaved from resin according to literature

procedures.[12]   UV (H_2O) λ_{max} 310 nm (68720).   MALDI-TOF-MS calcd. for

$C_{45}H_{60}N_{19}O_8$ (M + H): 1096.1. Found 1096.4.

ATK-iix-74. Synthesized on solid support and cleaved from resin according to literature

procedures.[12]   UV (H_2O) λ_{max} 310 nm (68720).   MALDI-TOF-MS calcd. for

$C_{45}H_{60}N_{19}O_8$ (M + H): 1096.1. Found 1096.3.

ATK-iix-75. Synthesized on solid support and cleaved from resin according to literature

procedures.[12]   UV (H_2O) λ_{max} 310 nm (68720).   MALDI-TOF-MS calcd. for

$C_{45}H_{60}N_{19}O_8$ (M + H): 1181.2. Found 1181.2.

ATK-iix-76. Synthesized on solid support and cleaved from resin according to literature

procedures.[12]   UV (H_2O) λ_{max} 310 nm (68720).   MALDI-TOF-MS calcd. for

$C_{45}H_{60}N_{19}O_8$ (M + H): 1181.2. Found 1181.6.

ATK-iix-77. Synthesized on solid support and cleaved from resin according to literature

procedures.[12]   UV (H_2O) λ_{max} 310 nm (68720).   MALDI-TOF-MS calcd. for

$C_{45}H_{60}N_{19}O_8$ (M + H): 1124.2. Found 1124.3.

Construction of plasmid DNA. Plasmid pATK4 was prepared by hybridization of complementary sets of synthetic oligonucleotides. The hybridized insert was individually ligated into *Bam*HI/*Hind*III-linearized pUC19 using T4 DNA ligase. *E. coli* JM109 high efficiency competent cells were then transformed with the ligated plasmid. Plasmid DNA from ampicillin-resistant white colonies was isolated using a Qiagen Wizard MidiPrep kit. The presence of the desired insert was determined by dideoxy sequencing. Concentration of prepared plasmid was determined by UV by the relationship 1 OD_{260} unit = 50 μg/mL duplex DNA.

Preparation of ^{32}P-end-labeled restriction fragments. Plasmid pATK4 was linearized with *Eco*RI and *Pvu*II restriction enzymes. The linearized plasmids were then treated with Klenow enzyme, deoxyadenosine 5'-[α-^{32}P]triphosphate, and thymidine 5'-[α-^{32}P]triphosphate for 3' labeling. The reactions were loaded onto a 7% nondenaturing polyacrylamide gel. The desired band was visualized by autoradiography and isolated. Chemical sequencing reactions were done according to published methods.

Quantitative DNase I footprinting.[13] DNase I footprinting reactions were carried out as previously described. Photostimulable storage phosphorimaging plates (Storage Phosphor Screen from Molecular Dynamics) were pressed flat against gel samples and exposed for 12–16 hours. Imaging of storage phosphor screens was accomplished on a Molecular Dynamics 425E PhosphorImager and the data analyzed using ImageQuant v. 3.2 software.

Binding energetics. Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl$_2$, 5 mM CaCl$_2$, pH 7.0, 22 °C) were performed on the 3'-^{32}P end labeled 270 bp *Eco*RI/*Pvu*II restriction fragment from pATK4. Equilibrium association constants for polyamides **ATK-iix-73 - 77** on the designed binding sites were determined by calculating a fractional saturation value at the site, for each polyamide concentration, and fitting the data to a modified Hill equation.

References.

1. Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology*, **2003,** *13*, 284–299.

2. Rucker, V. C.; Melander, C.; Dervan, P. B., *Helvetica Chimica Acta*, **2003,** *86*, 1839–1851.

3. Hawkins, C. A.; de Clairac, R. P.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society*, **2000,** *122*, 5235–5243.

4. Pelton, J. G.; Wemmer, D. E., *Journal of the American Chemical Society*, **1990,** *112*, 1393–1399.

5. Fagan, P.; Wemmer, D. E., *Journal of the American Chemical Society*, **1992,** *114*, 1080–1081.

6. Pelton, J. G.; Wemmer, D. E., *Proceedings of the National Academy of Sciences of the United States of America*, **1989,** *86*, 5723–5727.

7. Fagan, P.; *Ph. D. Thesis*; University of California Berkeley, Berkeley, **1996**.

8. Herman, D. M.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1998,** *120*, 1382–1391.

9. Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 2767–2774.

10. Swalley, S. E., **1996**, Unpublished Results.

11. Rucker, V. C.; Foister, S.; Melander, C.; Dervan, P. B., *Journal of the American Chemical Society*, **2003,** *125*, 1195–1202.

12. Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6141–6146.

13. Trauger, J. W.; Dervan, P. B., Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. In *Drug-Nucleic Acid Interactions*, 2001; Vol. 340, pp 450–466.

Appendix III

Imparting Sequence Specificity on Non-Specific DNA Enzymes with Hairpin Polyamide Conjugates

Abstract

Many of the enzymes responsible for maintaining and manipulating DNA do not possess inherent sequence specificity. Because of this, complexes of these enzymes with DNA substrates have been notoriously difficult to crystallize for structure determination. Because these structures would be of great use to the biochemical community, we report here initial efforts towards using DNA-binding polyamides to impart sequence specificity on these non-specific enzymes. We have synthesized a library of maleimide-conjugated polyamides for use in trapping cystine-labeled topoisomerase II. We also report the synthesis of chlorambucil-conjugated polyamides for use in trapping various members of the helicase family of enzymes. Biochemical and crystallographic studies are currently ongoing in the Berger lab at the University of California, Berkeley.

Introduction.

Many important enzymes and proteins contact DNA in a sequence-independent fashion. Chief among these are enzymes responsible for altering the topology of DNA during transcription. One example is topoisomerase II, which aids in solving the topological problems associated with DNA replication, transcription, and chromatin remodeling.[1] Topoisomerase II accomplishes this by binding duplex DNA, and creating
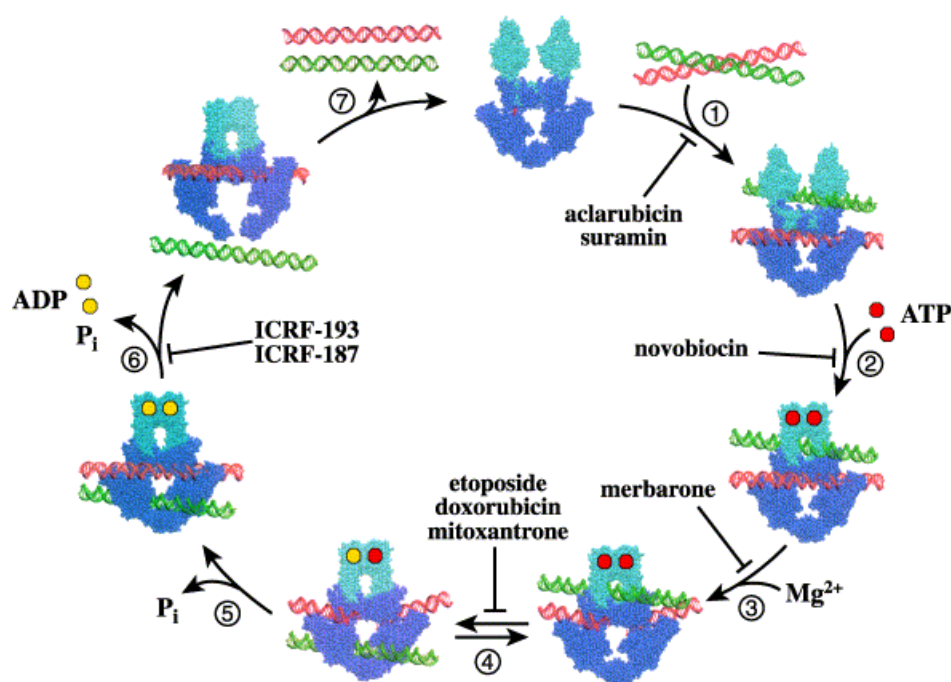

Figure III.1.[2] The catalytic cycle of topoisomerase II. 1 – The dimeric enzyme binds two strands of duplex DNA; 2, 3 – ATP is bound, and a double-strand break in one of the helices is formed; 4 – The intact strand is passed through the broken strand in an ATP-dependent fashion; 6, 7 – The cleaved strand is religated and the strands are released. Topoisomerase is an important cancer target, and many therapeutics have been developed to target various steps in this catalytic cycle (shown in the interior of the circle).

a double-strand break in the helix. The enzyme next passes a second double-helical region of DNA through this newly created break. Finally, the enzyme catalyzes the religation of the first helix (Figure III.1).[1, 3, 4] In this way, the enzyme helps relax DNA supercoils formed during transcription of circular DNA.

The helicase family of enzymes represents a second class of non-specific DNA-binding proteins. Helicases function as molecular motor proteins, moving unidirectionally along the double helix to unwind the energetically stable duplex DNA in an ATP-dependent fashion (Figure III.2).[5]
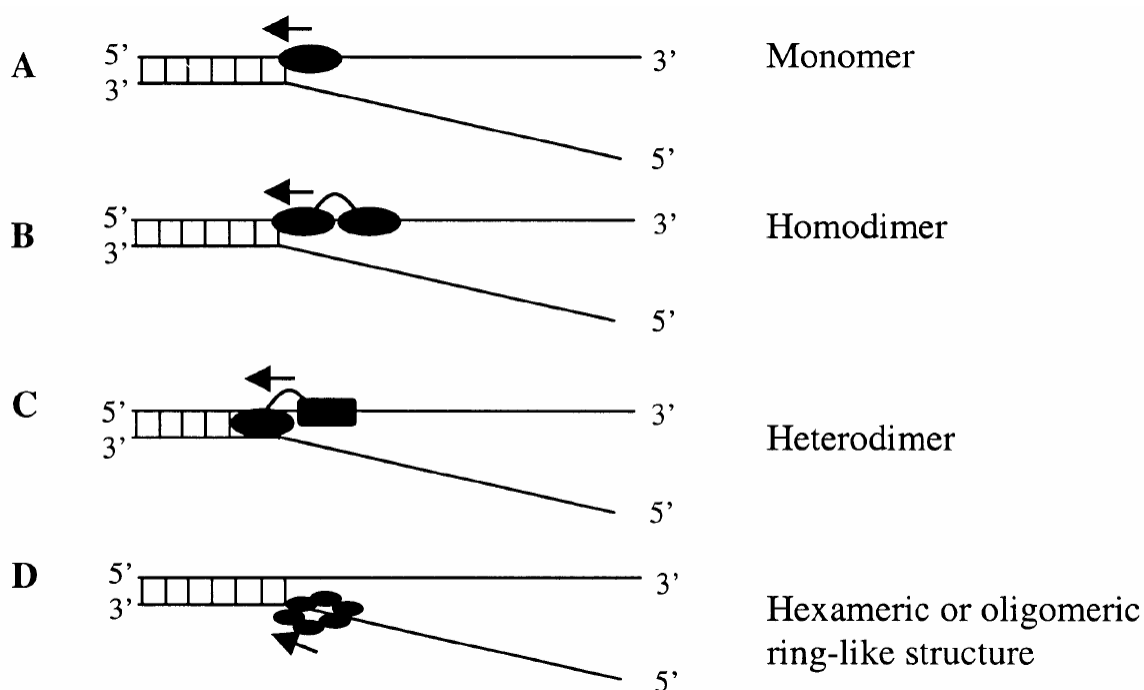

Figure III.2.[5] Schematic of the interaction of monomeric and oligomeric DNA helices with a forked DNA substrate. A – Monomeric helicase binds to both ssDNA and dsDNA. B – In homodimeric helicases, one subunit is always associated with the ssDNA along which it moves. C – Heterodimeric helicases contain a dsDNA-binding unit and an ssDNA-binding unit responsible for unwinding and translocation. D – Hexameric "ring-like" helicases encircle ssDNA, preventing local reannealing. One or more of the subunits binds the ssDNA/dsDNA junction.

Because topoisomerase II and helicase enzymes do not have any inherent sequence specificity, homogeneous crystals of the DNA:enzyme complexes have proven difficult to obtain. As a result, structural data on these complexes are currently unavailable.

Polyamides are small molecules capable of binding targeted sequences of DNA with high affinity and sequence specificity.[6] By conjugating protein-binding domains to

polyamides, researchers have been able to recruit proteins to targeted DNA sequences adjacent to binding sites for the polyamides.[7-10]   Perhaps polyamides can be used to impart some sequence specificity on non-specific DNA enzymes, thereby leading to homogeneous populations of protein:DNA:polyamide complexes suitable for crystal growth and structure determination.

Topoisomerase II.

As shown in Figure III.1, topoisomerase II is thought to bind DNA as a dimer, with the two arms of the dimer fully encircling the double helix of DNA. Normally, this binding can occur at any sequence (although GC-rich regions are slightly preferred). In our efforts towards homogeneous topoisomerase II:DNA complexes for crystallization, we use sequence-specific polyamides to target topoisomerase II to a discrete site on a duplex template. Our experimental design is illustrated in Figure III.3.

Briefly, a series of topoisomerase II mutants will be engineered to contain a single reactive functionality at various sites flanking the DNA-binding domain. Polyamide conjugates containing a complementary reactive group will be targeted to sites flanking an optimal topoisomerase binding site. When both topoisomerase II and polyamide are bound, a covalent linkage will be created, transferring the sequence specificity of the polyamide to the enzyme, and effectively trapping the topoisomerase II enzyme at a site adjacent to the polyamide binding site. Because the newlyimparted sequence specificity should yield homogeneous populations of complexes, crystals will be grown, and structures determined.
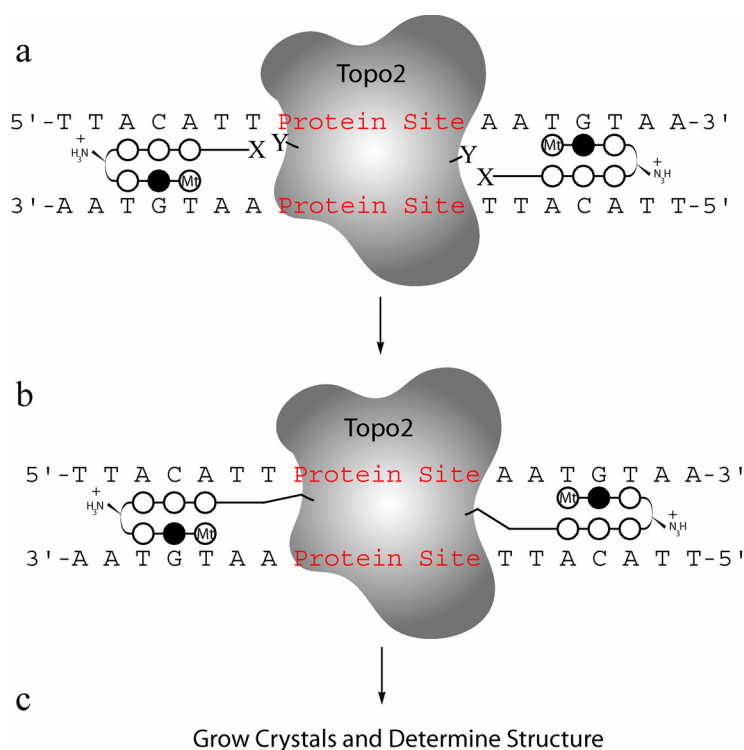
Figure III.3. Schematic representation of polyamide-mediated trapping of topoisomerase II. a – Polyamides functionalized with X and topoisomerase II functionalized with Y (where X and Y are complementary reactive groups) bind DNA at adjacent sites. b – Due to proximity, X and Y react, forming a covalent link that imparts the sequence specificity of the polyamide onto the topoisomerase II enzyme. c – Homogeneous complexes are then used for crystal growth and structure determination.

We have chosen a Michael addition between a cystine thiol residue on the protein and a maleimide functionality on the polyamide for initial attachment reactions. A series of mutant topoisomerase II enzymes will be created in which a single cystine mutation will be made at various residues flanking the DNA-binding domain (Figure III.5). Once these mutants have been tested to ensure the retention of their DNA-binding properties and function, they will be incubated with a DNA template and maleimide-functionalized polyamides. We chose to use the sequences 5'-CCG GT<u>T ACA T</u>T(G **GCC**)$_n$ **GAT CGG CCG ATC** (**GGC** C)$_n$A<u>A TGT A</u>AC CGG-3' for our studies. These sequences are palindromic, and contain a GC-rich domain for topoisomerase binding (bold), flanked by two polyamide binding sites (underlined). Because the precise distance between the
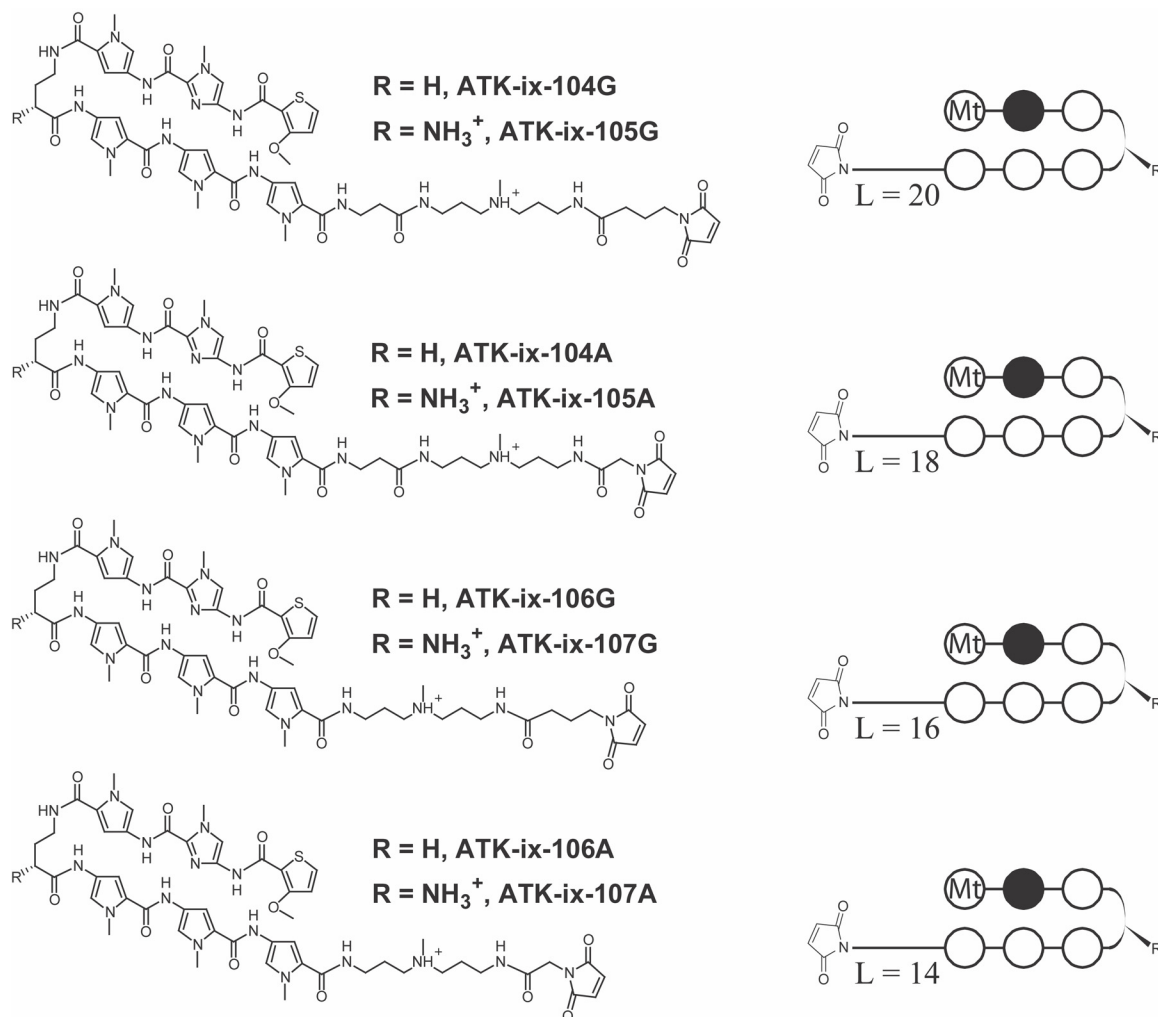
Figure III.4. Chemical and schematic structures of the polyamide-maleimide conjugates synthesized for topoisomerase II trapping. Distances from the DNA-binding portion to the reactive maleimide range from 14–20 atoms. Analogs of each compound were synthesized with and without the chiral turn (to modulate DNA-binding affinity). Shown below each schematic is the linker length in atoms (L = n).

polyamide and topoisomerase II binding sites optimal for conjugate addition is not known, several templates (n = 0, 1, 2, and 3) will be tested. The polyamides synthesized for thus study are illustrated in Figure III.4. We chose a polyamide directed towards the site 5'-WTGWW-3' (where W = A or T) because of the sequence's high AT content. This polyamide should thus not bind the GC-rich topoisomerase-binding region. The
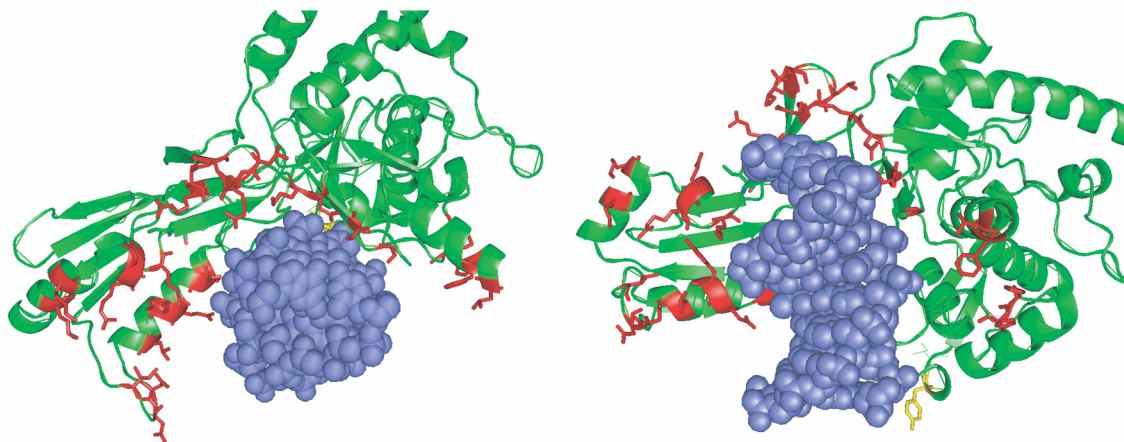
Figure III.5. Two views of a molecular model of the DNA-binding portion of topoisomerase II (green ribbons) complexed with duplex DNA (blue spheres). Highlighted in red are the residues chosen for cysteine mutations. Each mutant topoisomerase II will contain a single cysteine substitution at one of the residues highlighted in red. All residues chosen lay within 10 Å of the modeled DNA helix.

polyamide library consists of molecules with an identical polyamide core functionalized with several different linker lengths connecting the DNA-binding polyamide to the cystine-reactive maleimide. Analogs with and without the chiral turn were synthesized to give compounds with weak (no chiral turn) and strong (chiral turn) binding affinities. Compounds were sent to the Berger lab at the University of California, Berkeley, in November 2004.

Helicase.

As shown in Figure III.2, helicases slide non-specifically along the double helix, melting the two strands as they progress. Polyamide-chlorambucil conjugates have been shown to create interstrand DNA crosslinks in a sequence-specific fashion (Figure III.6).[11, 12] In an effort to create homogeneous DNA:helicase complexes for crystallographic structure determination, we propose the use of polyamide-chlorambucil conjugates to act as molecular "chocks" to stop helicase progression at a predetermined site (Figure III.7). In such a scheme, polyamide-chlorambucil conjugates will be
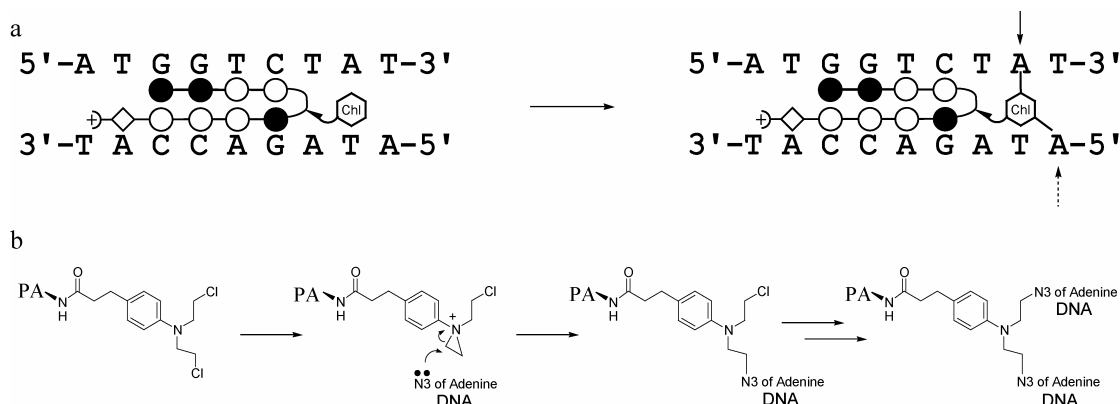
a

b

Figure III.6.   a.   Schematic of sequence-specific DNA crosslinking by a polyamide-chlorambucil conjugate.[12]   At right, dark arrow indicates major alkylation site; dashed arrow indicates minor alkylation site.   b.   Mechanism of action of chlorambucil.   The lone pair of electrons on N3 of adenine attacks the aziridine, forming a covalent bond.   Because there are two chloroethyl functionalities, this reaction can occur twice to create an interstrand crosslink.
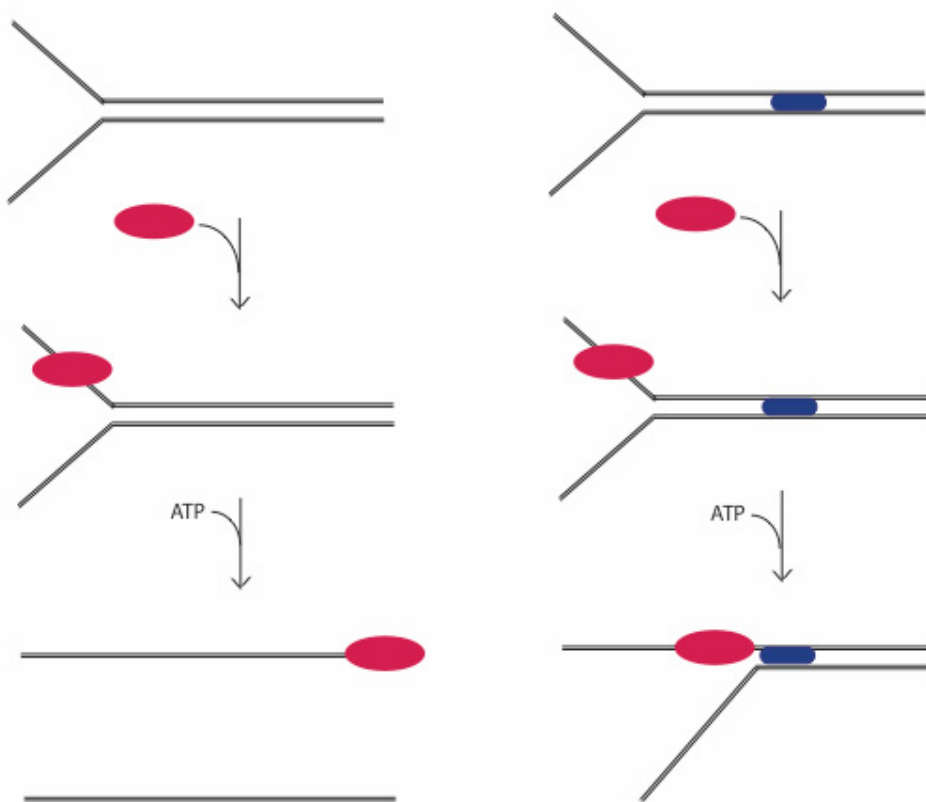

Figure III.7.   Left column: Schematic of helicase activity on DNA.   Helicase (red oval) binds to the ssDNA region of a fork template.   The helicase then translocates in an ATP-dependent fashion to unwind the DNA.   Right Column: Polyamide-chlorambucil conjugates will bind to a designed sequence in the dsDNA region of the fork template.   The nitrogen mustard agent will then form an interstrand crosslink. Helicase will be added.   The DNA will be unwound until the enzyme reaches the polyamide "chock," where it will be unable to continue.   Homogeneous complexes will then by used for crystal growth and structure determination.

incubated with a DNA template for helicase unwinding. The two DNA strands will be crosslinked at a single site adjacent to the polyamide binding site. Helicase enzymes will then be allowed to progress towards the polyamide "stopper." Upon reaching the crosslink site, progression will be arrested. Because progression is arrested at a discrete site (dependent upon polyamide binding), homogeneous populations of complexes should result, which will then be used for structure determination.
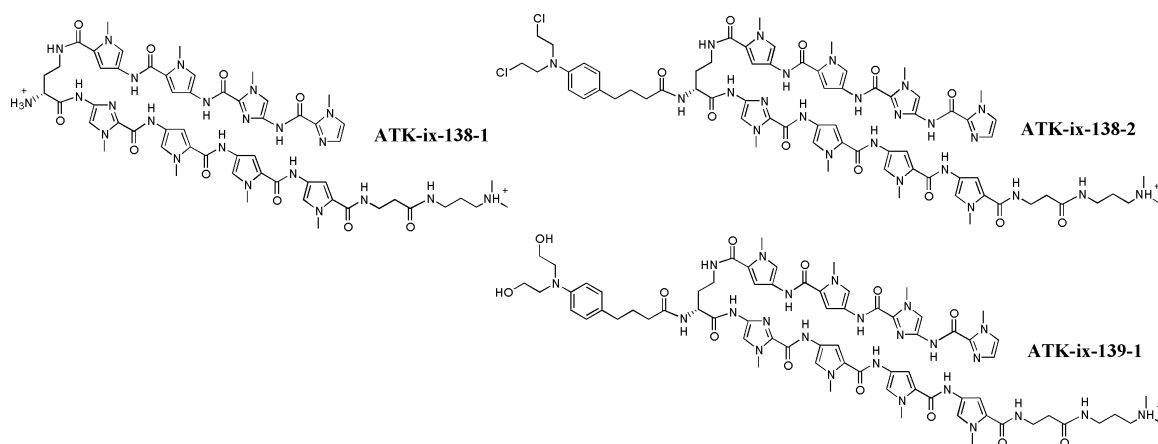

Figure III.8. Chemical structures of the polyamides synthesized for helicase trapping studies. Compound **ATK-iix-138-2** is the chlorambucil-containing polyamide. **ATK-iix-138-1** and **ATK-iix-139-1** are negative control compounds where the chlorambucil has been removed and hydrolyzed, respectively.

Because the DNA construct for this study is artificial, we chose to use the polyamide core ImImPyPy-γ-ImImPyPy-β-Dp for crosslinking studies. This polyamide has previously proven to bind DNA with high affinity and excellent specificity.[13] Because our main concern is homogeneity in the site of helicase arrest, this highly specific polyamide was chosen. Polyamide conjugates shown in Figure III.8 have been synthesized and sent to the Berger Lab, at the University of California, Berkeley, as of

April 2005. The polyamides will be tested against a library of helicases, including MCM, DnaB, and T7 phage helicase.

Conclusion.

Polyamide conjugates have been designed and synthesized to impart sequence specificity onto non-specific DNA enzymes in order to create homogeneous complexes that are able to be crystallized for structural determination. Topoisomerase II and helicase enzymes are the targets of the present research. Because no precise structures are currently available with these enzymes in complex with their DNA substrates, they would be of great use to the field.

Materials and Methods.

Polyamides were synthesized according to standard procedures on either PAM or oxime resin.[14, 15] Polyamides **ATK-iix-104 – 107** were synthesized using the methoxythiophene cap as previously reported.[16]

Sample procedure for the synthesis of maleimide conjugates.

Base polyamides were cleaved from either oxime or PAM resin with methylamino dipropylamine and purified by preparative HPLC. 1 μmol polyamide amine was combined in 100 μL 10:1 DMF:DIEA with 5 μM of maleimide NHS esters (Pierce). The reaction was monitored by analytical HPLC, and when done, purified by HPLC. Typical yields after resin cleavage were ~ 70%.

ATK-iix-104A UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1191.3. Found 1191.3.

ATK-iix-104G UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1219.4 Found 1219.3.

ATK-iix-105A UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1207.3 Found 1207.8.

ATK-iix-105G UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1235.4. Found 1235.5.

ATK-iix-106A UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1120.2. Found 1120.1.

ATK-iix-106G UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1148.3. Found 1148.1.

ATK-iix-107A UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1136.2. Found 1136.2.

ATK-iix-107G UV (H$_2$O) λ_{max} 310 nm (51540). MALDI-TOF-MS calcd. (M + H): 1164.3. Found 1164.6.

ATK-iix-138-2

Polyamide **ATK-iix-138-1** was liberated from resin by aminolysis in neat dimethylamino propylamine (Dp) for 12 hours at 50 ºC and purified by preparative HPLC. **ATK-iix-138-1** (2 μmol) was dissolved in 300 μL DMF and 10 μL DIEA. Chlorambucil (4 μmol, 1.33 mg) was added, followed quickly by PyBOP (3.9 μmol, 2.04 mg). The reaction was mixed, and allowed to stand at room temperature for 45 minutes. The reaction was then diluted with 0.1% TFA in water and purified by reversed-phase HPLC. To avoid hydrolysis, as soon as the product eluted from the column, 500 μL aliquots were freeze-dried. Product was isolated in 50% yield. UV (H_2O) λ_{max} 310 nm (68720). MALDI-TOF-MS calcd. (M + H): 1525.5. Found 1525.7.

ATK-iix-139-1

Polyamide **ATK-iix-138-2** (500 nmol) was dissolved in 100 μL DMF. 1 mL of 25% NaOH in water was added, and the reaction mixed at room temperature for four hours. The reaction was acidified with TFA, diluted to 10 mL with water, and purified by reversed-phase HPLC (200 nmol yield).

References.

1. Champoux, J. J., *Annual Review of Biochemistry*, **2001,** *70*, 369–413.

2. Larsen, A. K.; Eseargueil, A. E.; Skladanowski, A., *Pharmacology & Therapeutics*, **2003,** *99*, 167–181.

3. Baird, C. L.; Gordon, M. S.; Andrenyak, D. M.; Marecek, J. F.; Lindsley, J. E., *Journal of Biological Chemistry*, **2001,** *276*, 27893–27898.

4. Fass, D.; Bogden, C. E.; Berger, J. M., *Nature Structural Biology*, **1999,** *6*, 322–326.

5. Tuteja, N.; Tuteja, R., *European Journal of Biochemistry*, **2004,** *271*, 1849–1863.

6. Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology*, **2003,** *13*, 284–299.

7. Ansari, A. Z.; Mapp, A. K.; Nguyen, D. H.; Dervan, P. B.; Ptashne, M., *Chemistry & Biology*, **2001,** *8*, 583–592.

8. Arndt, H. D.; Hauschild, K. E.; Sullivan, D. P.; Lake, K.; Dervan, P. B.; Ansari, A. Z., *Journal of the American Chemical Society*, **2003,** *125*, 13322–13323.

9. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America*, **2000,** *97*, 3930–3935.

10. Wang, C. C. C.; Dervan, P. B., *Journal of the American Chemical Society*, **2001,** *123*, 8657–8661.

11. Wang, Y. D.; Dziegielewski, J.; Wurtz, N. R.; Dziegielewska, B.; Dervan, P. B.; Beerman, T. A., *Nucleic Acids Research*, **2003,** *31*, 1208–1215.

12. Wurtz, N. R.; Dervan, P. B., *Chemistry & Biology*, **2000,** *7*, 153–161.

13. Poulin-Kerstien, A. T. DNA-Templated Dimerizations of Minor Groove-Binding Polyamides. California Institute of Technology, Pasadena, **2005**.

14. Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society*, **1996,** *118*, 6141–6146.

15. Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2002,** *10*, 2767–2774.

16. Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., *Bioorganic & Medicinal Chemistry*, **2003,** *11*, 4333–4340.