Chapter 7

Programmable Oligomers for DNA Recognition

Abstract.

Hairpin polyamides have been shown to selectively recognize predetermined DNA sequences with high affinities and sequence specificities. Polyamides are able to discriminate between each of the four Watson-Crick Base pairs through the side-by-side pairing of aromatic rings in the minor groove of DNA. While the established polyamide/DNA pairing rules have been effective at targeting hundreds of sequences *in vivo* and *in vitro*, we herein present a step towards designing polyamide-based oligomers which contain an emerging set of recognition elements capable of targeting a new selection of DNA sequences. The first of these elements includes the N-terminal **No-Hz** and **Ct-Hz** dimer caps, which are able to target G-T and T-T sequences, respectively, with good affinities and specificities.

Introduction.

Hairpin polyamides represent a growing class of sequence-specific, DNA binding molecules.[3-5]   First-generation polyamides comprising the heterocyclic amino acid subunits 1-methyl-1*H*-imidazole (**Im**), 1-methyl-1*H*-pyrrole (**Py**) and 3-hydroxy-1*H*-pyrrole (**Hp**) are able to bind in the minor groove of DNA with sub-nanomolar affinities and significant sequence selectivity. While co-facial pairings of **Im**, **Py**, and **Hp** where Im/Py is specific for G•C, Hp/Py is specific for T•A and Py/Py is specific for A•T and T•A, are able to code for the four Watson-Crick base pairs,[6-8] the high level of structural variation between DNA sequences makes it especially difficult for such a small repertoire of recognition elements to effectively target a diverse range of sequences. Even with substantial, sequence-specific differences in the DNA microstructure, the "pairing rules" have proven effective in recognizing hundreds of predetermined and biological DNA targets.[9, 10]

Much inspiration for the structural design of standard **Im**, **Py**, and **Hp** polyamides was drawn from the natural products netropsin and dystamycin A.[11] Comprised of amide linked pyrrole rings, these small molecules are capable of selectively binding A,T tracks of DNA in the minor groove. Subsequent, deliberate atomic changes to the edges of the antiparallel aromatic amino acid ring pairs that interact with the target DNA led to the advent of the G specific **Im** residue and the T-specific **Hp** residue.[12, 13] Placing an aromatic nitrogen in the minor groove allowed **Im** to accept a hydrogen bond from the exocyclic amine of guanine, thus imparting excellent G > C specificity. Incorporating an exocyclic hydroxyl group in **Hp** allows the ring to specify for T > A; the hydroxyl groups

prefer to lie across the less bulky thymine base in an A•T base pair while also forming hydrogen bonds with the O2 carbonyl of thymine.

Our initial efforts to further expand the polyamide alphabet utilized the same approach of single atom substitutions on the existing 5-membered heterocycle scaffold.[14] We believed that evaluating many heterocycles, each with relatively subtle structural differences, would lead us to a ring pairing that could serve as a novel mode of DNA recognition. To our surprise, an extensive study of many new ring systems led only to a single result wherein we found that a thiophene ring (**Tn**) was able to present a sulfur atom to the minor groove thus imparting a modest selectivity for T > A in a T•A base pair.

It was not until we explored the properties of the thiophene rings at the polyamide N-terminal (cap) position that the full significance of the **Tn** result became apparent: internal pairings and cap pairings were not equivalent.[15] The notion that ring pairings were context dependent led to a veritable paradigm shift in our approach to polyamide design (Figure 7.1). Ring
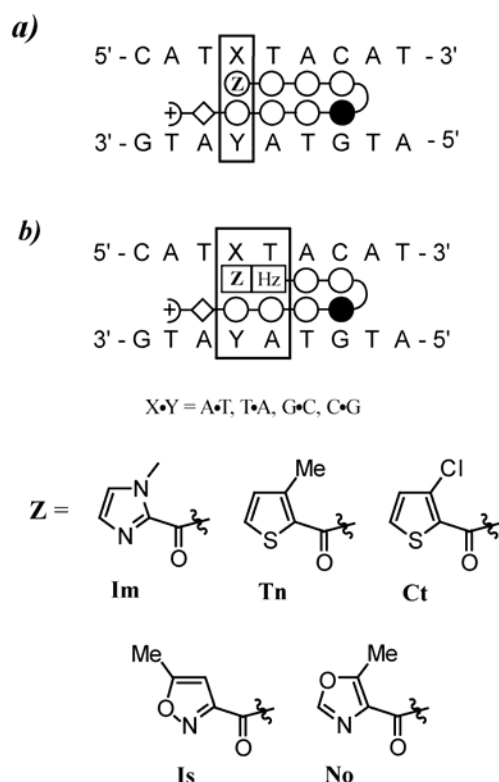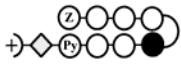

Figure 7.1. Experimental scheme showing the DNA sequences that each polyamide will be tested against. (a) Hairpin polyamides containing novel rings (Z = Im, Tn, Ct, Is, No) proximal to a pyrrole (**Py**) residue. (b) Hairpin polyamides containing novel rings (Z) proximal to a hydroxybenzimidazole residue (**Hz**). Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

pairings that exhibited poor binding properties and little or no specificity internally, such as the (chlorothiophene) **Ct/Py** pair, were much more successful at inferring an appreciable level of specificity at the cap position (Table 7.1).

Table 7.1. Affinities of Z/Py Ring Pairs K_a [M^{-1}][a,b]

	A•T	T•A	G•C	C•G
Im/Py	2.8 (±0.2)x 10^9	3.8 (±0.3)x 10^9	7.0 (±0.9)x 10^{10}	3.2 (±0.4)x 10^9
Tn/Py	1.4 (±0.2)x 10^9	2.3 (±0.4)x 10^9	1.0 (±0.4)x 10^7	1.0 (±0.3)x 10^7
Ct/Py	3.7 (±0.2)x 10^9	1.3 (±0.2)x 10^{10}	3.1 (±0.6)x 10^8	2.1 (±1.1)x 10^8
Is/Py	3.2 (±0.3)x 10^9	1.1 (±0.5)x 10^9	1.3 (±0.3)x 10^9	1.9 (±0.8)x 10^9
No/Py	3.4 (±0.3)x 10^9	1.0 (±0.6)x 10^{10}	4.1 (±0.3)x 10^{10}	7.5 (±0.6)x 10^9

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

We herein report a new set of heterocyclic dimers that represent a step away from single base-pair recognition and towards a more global approach to molecular recognition. The new recognition elements were designed by combining the T-specific hydroxybenzimidazole 6-5 fused ring system (**Hz**) at the internal position with the newly developed thiophene ring caps (Figure 7.2).[16, 17] When used at the cap position of hairpin polyamides, the **Ct-Hz** and **No-Hz** dimers are able to target T-T and G-T sequences, respectively, with high affinity and appreciable specificity. Evolution of the **No-Hz** dimer came about due to the inability of **Im-Hz** to specify for its designed G-T site. DNase I footprinting titrations were used to determine DNA binding affinities of the **Ct-Hz** and **No-Hz** dimers for the four Watson-Crick bases. Molecular modeling was used to further complement the thermodynamic results.
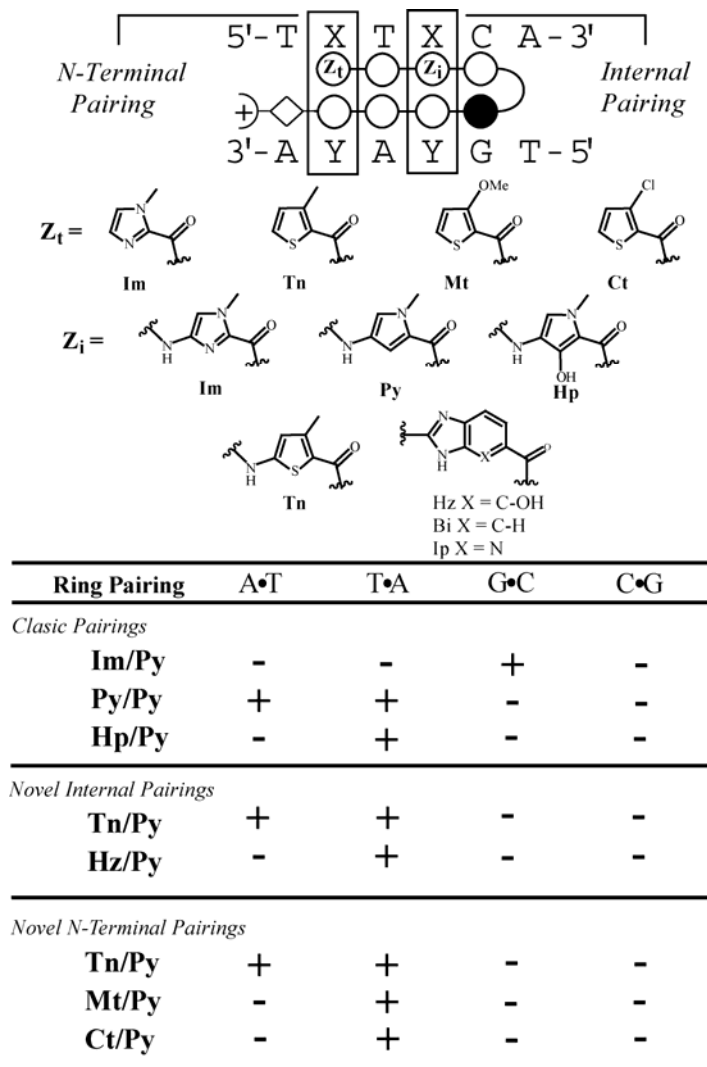
Figure 7.2. Classic and novel heterocycles used for DNA recognition. (Above) Heterocycles used as recognition elements at the N-terminus and internal positions of hairpin polyamides. Zt and Zi, respectively. (Below) Specificites of the classic imidazole (**Im**), pyrrole (**Py**) and hydroxypyrrole (**Hp**) ring-ring pairings, as well as specificities for recently developed internal and N-terminus pairs.

Ring Pairing	A•T	T•A	G•C	C•G
Clasic Pairings				
Im/Py	-	-	+	-
Py/Py	+	+	-	-
Hp/Py	-	+	-	-
Novel Internal Pairings				
Tn/Py	+	+	-	-
Hz/Py	-	+	-	-
Novel N-Terminal Pairings				
Tn/Py	+	+	-	-
Mt/Py	-	+	-	-
Ct/Py	-	+	-	-

Results.

Heterocycle Synthesis. No-OH (7). To a mixture of ethyl-alpha-isocyanoacetate and DBU in dry THF at 10 °C was added a solution of acetic anhydride in dry THF dropwise. After warming to R.T. the reaction was allowed to stir for 10 hours at which point the solvent was removed by rotoevaporation, water was added and the product extracted with EtOAc. The organic layer was dried over sodium sulfate and removed by rotoevaporation to yield **No-OMe** (**6**) as a crude amber oil, which, after column chromatography, provided the ester as a crystalline white solid in moderate yield. The ester (**6**) was then saponified using sodium hydroxide in methanol at elevated temperature. After rotoevaporation of the methanol solvent, the pH of the aqueous solution was adjusted and resulting precipitate extracted with EtOAc, dried over
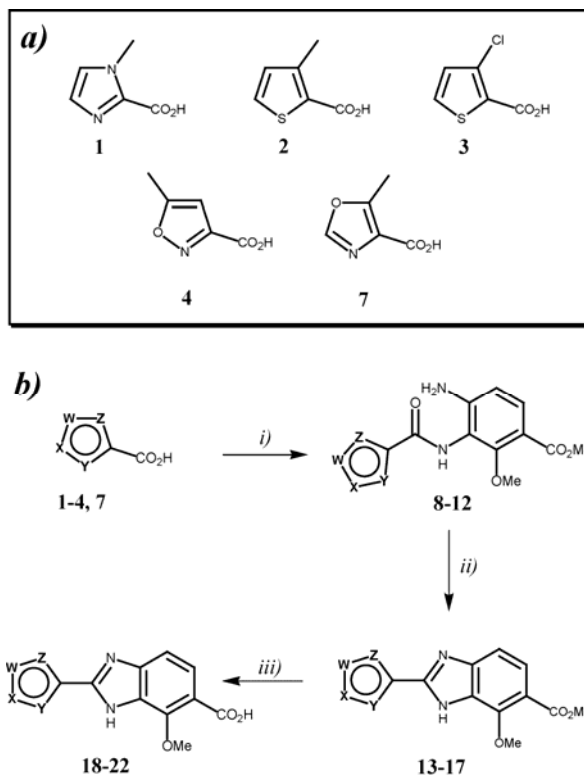
Figure 7.3. Synthesis of heterocyclic amino acid monomers and dimers. (a) 5-membered heterocyclic caps **1-4** and **7**. General procedure for the synthesis of fused heterocyclic dimers: (i) Heterocyclic Acids, HBTU, DIEA, DMF, aryl diamine (**33**) 35 ^{o}C, 24 h; (ii) AcOH, 90 ^{o}C, 12 h; (iii) NaOH, MeOH, 60 ^{o}C, 4-6 h. (**1**, **8**, **13**, **18**) W = C-H, X = C-H, Y = N, Z = N-Me; (**2**, **9**, **14**, **19**) W = C-H, X = C-H, Y = S, Z = C-Me; (**3**, **10**, **15**, **20**) W = C-H, X = C-H, Y = S, Z = C-Cl; (**4**, **11**, **16**, **21**) W = C-Me, X = O, Y = N, Z = C-H; (**7**, **12**, **17**, **22**) W = O, X = C-H, Y = N, Z = C-Me.

sodium sulfate and concentrated in vacuo to provide **No-OH** (**7**) as a fibrous white solid in good yield.

General Synthetic Procedure Fused Bicycle Formation. (Dimeric caps **18-22**) To a solution of heterocyclic carboxylic acids **1-4**, and **7** in DMF was added DIEA and HBTU. The solution was allowed to stir at R.T. for 1 hour, allowing for conversion of the carboxylic acid to the activated aryl HBTU-ester. The aryl diamine (**35**) was then added and the reaction heated to 35 ^{o}C for 24 hours. The reaction was allowed to cool to room temperature and poured into a separatory funnel containing water. The water was then extracted with EtOAc, dried over sodium sulfate, and removed by rotoevaporation to provide the amides as crude solids. The amides were then dissolved in acetic acid and heated to 90°C for 12 hours, at which point the AcOH was removed by rotoevaporation. The resultant residues were subjected to column chromatography (EtOAc/Hex) to provide the dimeric methyl ester caps as a thin films.
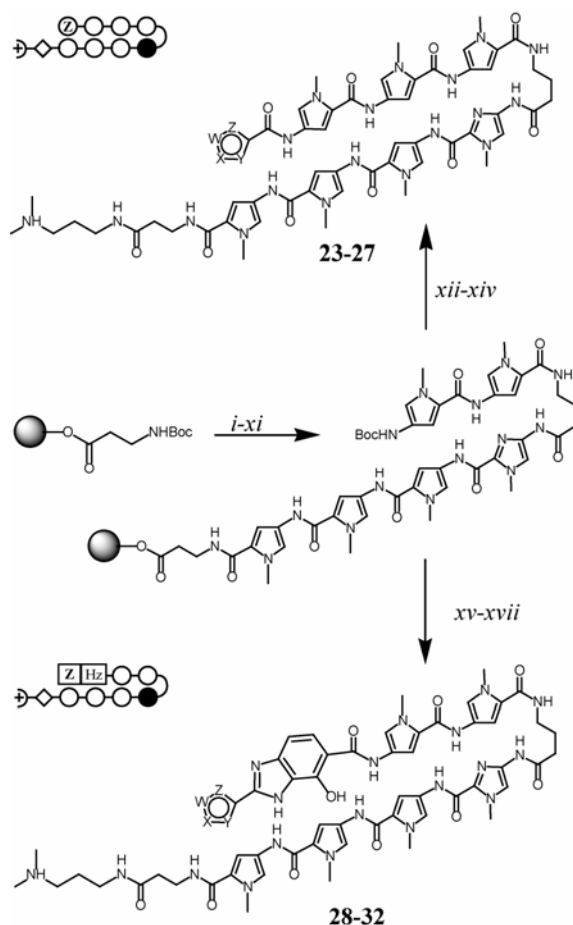
Figure 7.4. Solid-phase synthesis of hairpin oligomers. i) TFA/DCM; ii) Boc-Py-OBt, DIEA, DMF; iii) Ac$_2$O, DIEA, DMF; iv) Steps i-iii (x2); v) TFA/DCM; vi) Boc-Im-OH, HBTU, DIEA, DMF; vii) Ac$_2$O, DIEA, DMF; viii) TFA/DCM; ix) Boc-γ-OH, HBTU, DIEA, DMF; x) Ac$_2$O, DIEA, DMF; xi) Steps i-iii (x2); xii) Steps i-iii (x1); xiii) Terminal Heterocyclic Acid, HBTU, DIEA, DMF; xiv) Dp, 80 °C, 2h; xv) TFA/DCM; xvi) Terminal Heterocyclic Dimer, HBTU, DIEA, DMF; xvii) Dp, 80 °C, 2h. Polyamides: (**23, 28**) W = C-H, X = C-H, Y = N, Z = N-Me; (**24, 29**) W = C-H, X = C-H, Y = S, Z = C-Me; (**25, 30**) W = C-H, X = C-H, Y = S, Z = C-Cl; (**26, 31**) W = C-Me, X = O, Y = N, Z = C-H; (**27, 32**) W = O, X = C-H, Y = N, Z = C-Me.

Addition of hexanes, followed by rotoevaporation and drying under high vacuum provided the dimers as white solids. The dimeric esters were then saponified using 1N NaOH and MeOH at 40 °C for 4 h. Neutralization and collection of the precipitate provided the dimeric cap compounds **18**-**22** (Figure 7.3).

Oligomer Synthesis. Hairpin polyamides were synthesized manually from Boc-β-PAM resin in a stepwise fashion using Boc-protected monomeric and dimeric amino acids according to established solid-phase protocols.[18] Base Resin 1 (**BR1**) (H$_2$N-Py-Py-γ-Im-Py-Py-Py-β-Pam) and Base Resin 2 (**BR2**) (H$_2$N-Py-Py-Py-γ-Im-Py-Py-Py-β-Pam) were synthesized in gram quantities using the following amino acid building blocks: Boc-**Py**-OBt, Boc-**Im**-OH,

and Boc-γ-OH (Figure 7.4). The base resins were then split into smaller batches for coupling to the final monomeric and dimeric caps. Boc-protected amino acid monomers and dimers for **Im**, **Py**, and **Ct** were synthesized according to previously reported procedures.[15, 18, 19] Couplings were realized using pre-activated monomers (Boc-**Py**-OBt) or HBTU activation in a DIEA and DMF mixture. Coupling times ran from 3-24 h at 25-40 $^{\circ}$C. Deprotection of the growing polyamide was accomplished using 80% TFA/DCM. Polyamides were cleaved from the resin by treatment with dimethylaminopropylamine (Dp) neat at 80 $^{\circ}$C for 2 h, and purified by preparatory reverse phase HPLC. Deprotection of the methoxy-protected polyamides was done using a mix of thiophenoxide in DMF at 80 $^{\circ}$C, to provide the following free hydroxy derivatives after a second HPLC purification: **Im**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**23**), **Tn**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**24**), **Ct**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**25**), **Is**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**26**), **No**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**27**), **Im**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**28**), **Tn**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**29**), **Ct**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**30**), **Is**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**31**), **No**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**32**).

DNA Affinity and Sequence Specificity. Quantitative DNase-I footprint titrations were carried out for oligomers **23-32**. All oligomers were footprinted on the 285-base-pair PCR product of plasmid pCW15. In all cases, the DNA sequence specificity at cap position (in bold) was determined by varying a single DNA base pair within the sequence, 5'-T**X**TACA-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes. The variable base-pair position was designed to be adjacent to the **Hz** ring, which has been shown to specify
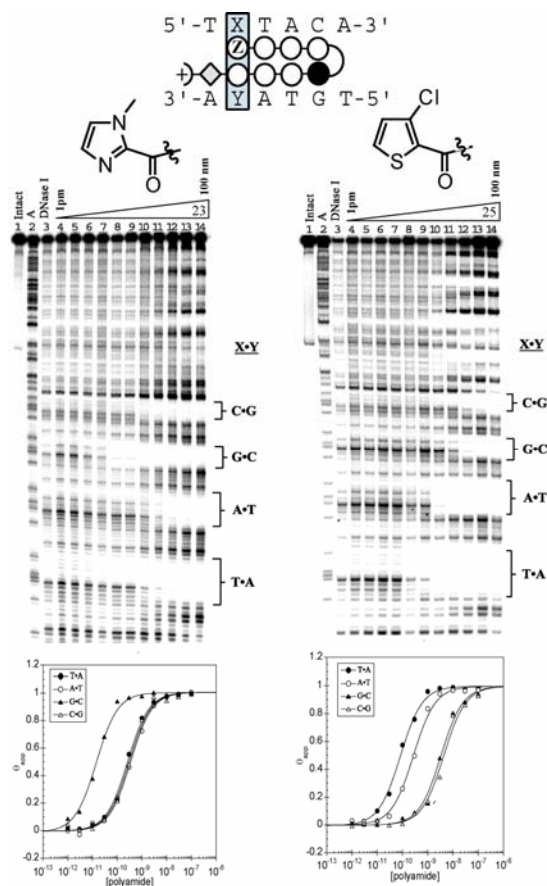
Figure 7.5. Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **23** and **25**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid CW15: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (top) Chemical structure of the pairing of interest; and (bottom) binding isotherms for the four designed sites. θ norm values were obtained according to published methods.[1] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

for T when paired across from **Py**, so as to be able to determine the binding properties of each compound to the following two base-pair *sequences*: **AT**, **TT**, **GT** and **CT**.

Hairpins **Im**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**23**) and **Ct**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**25**) serve as the parent compounds from which the dimer-cap-containing oligomers are derived. As shown by their binding profiles (Figure 7.5), both systems are well characterized with the **Im** cap hairpin exhibiting specificity for G at the cap position and the **Ct** cap hairpin exhibiting specificity for T (Table 7.1).

The sequence specificity of the **Ct-Hz** and **Im-Hz** dimers for 5'-T**X**TACA-3' was evaluated using oligomers **30** and **28** respectively. As expected polyamide **30** bound its designed match site 5'-T**TT**ACA-3' (K_a = 2.4 x 10^9) (Figure 7.6) with both the **Ct** and the **Hz** halves of the dimer preferring to
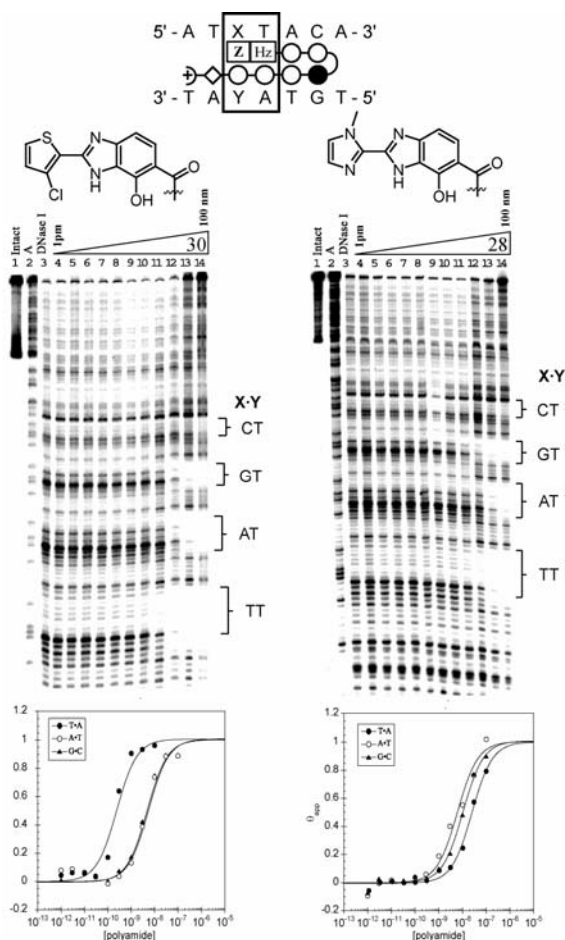
Figure 7.6. Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **30** and **28**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid CW15: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (top) Chemical structure of the pairing of interest; and (bottom) binding isotherms for the four designed sites. θ norm values were obtained according to published methods.[1] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

rest over the less bulky T in a T•A base pair. To our delight, placing the **Ct** ring adjacent to the **Hz** seemingly increased the specificity of the thiophene ring for T > A from ~4-fold in a **Ct**-Py system to 10-fold on the **Ct-Hz** system. Polyamide **28**, which contains the **Im-Hz** dimer, did not bind its designed match site 5'-T**GT**ACA-3' with any appreciable level of specificity exhibiting affinities of (K_a = 1.6 x 10^8) and (K_a = 4.0 x 10^8) for the **GT** and **AT** sites respectively.

The binding properties of **Is** and **No** were accessed in the very well characterized 8-ring hairpin system as a means of pre-screening them for any G•C over C•G specificity and for eventual use as **Im** replacements in the dimer cap system (Figure 7.7). Oligomer **26**, which contains the isoxazole end-cap showed a complete lack of specificity for any of the
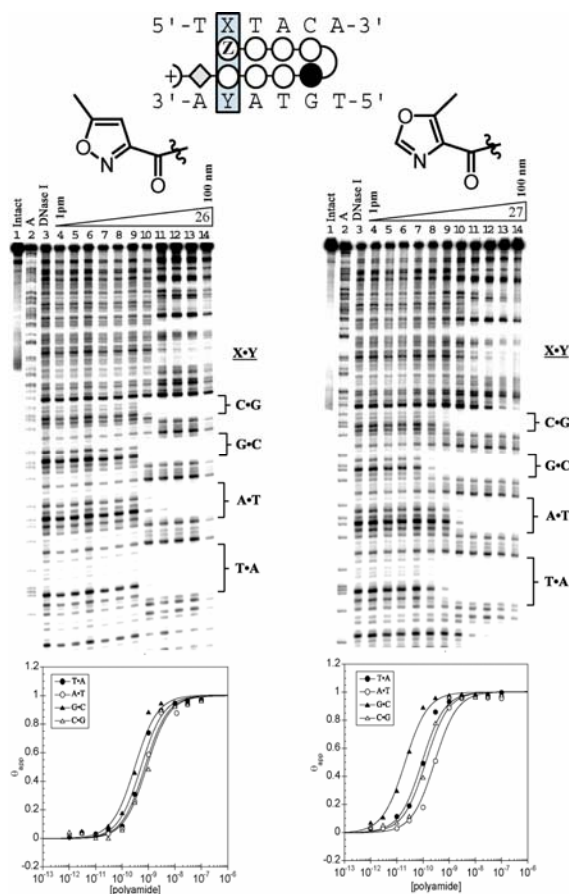
Figure 7.7. Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **26** and **27**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid CW15: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (top) Chemical structure of the pairing of interest; and (bottom) binding isotherms for the four designed sites. θ norm values were obtained according to published methods.[1] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

Watson-Crick base pairs. Oligomer **27** which contains the oxazole end-cap, however, displayed a 5.5-fold preference for G > C (Table 7.1).

Both the isoxazole (**31**) and oxazole caps (**32**) were incorporated into the dimer cap system and their affinities for their designed match site, 5'-**TGT**ACA-3', accessed. **31** was unable to discriminate between the **AT**, **TT** and **GT** sites with affinities of 1.1, 1.0 and 2.0×10^9 M^{-1} respectively. However, **34** successfully targeted it designed match site with an appreciable level of specificity (25-fold) and a match site affinity of $K_a = 6.8 \times 10^9$ M^{-1} (Figure 7.8 & Table 7.2).

Molecular Modeling. Modeling calculations were preformed using the *'Spartan Essential'* software package. All *Ab initio* calculations were conducted using the Hartree-Fock model
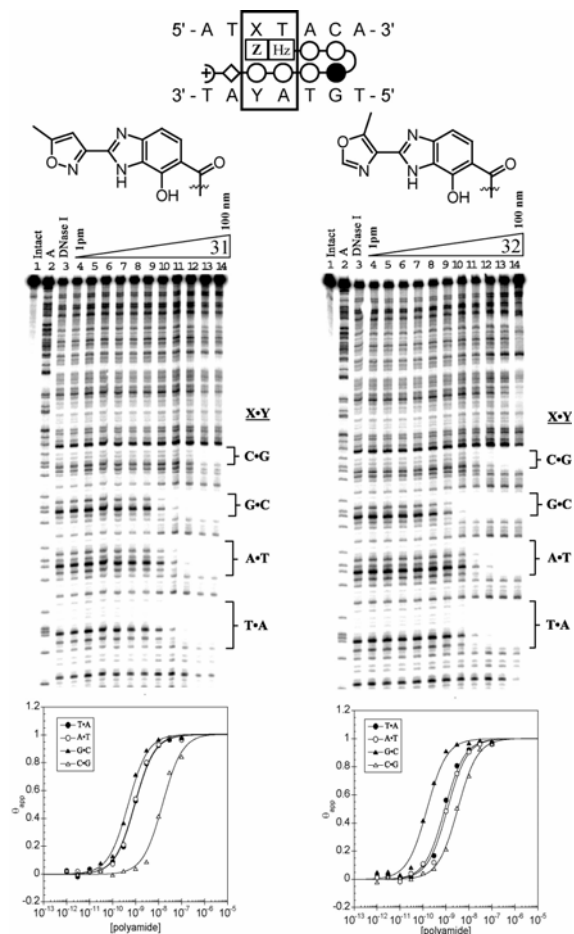
Figure 7.8. Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **31** and **32**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid CW15: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (top) Chemical structure of the pairing of interest; and (bottom) binding isotherms for the four designed sites. θ norm values were obtained according to published methods.[1] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

and a 6-31G* polarization basis set. Two dimers, **Im-Hz** and **Im-Py**, were modeled to examine the differences in polyamide curvature and proximity to the DNA bases and to establish a rationale for the decreased ability of Im to bind G > C when adjacent to the 6-5 fused hydroxybenzimidazole ring (Figure 7.9).

Discussion.

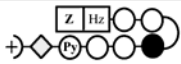Recent advances in hairpin polyamide designs have traditionally focused on developing new modes of single base pair recognition by heterocyclic ring pairings. Our studies, however, have highlighted the fact that the microstructure of DNA highly depends on the sequence in question.[20] Thus we have taken a more global view of molecular recognition where our efforts have shifted from designing molecules that

target DNA base pairs to designing those that target DNA sequences.

The **Ct-Hz** dimer cap represents our first effort to target a short sequence of DNA using sequence-inspired recognition elements. Previous efforts to target two consecutive thymine bases using adjacent **Hp** rings yielded compounds with poor affinities and no specificity. Studies have also shown that **Hz** exhibits specificity for T > A at the N-1 position – relative to the polyamide N-terminus – and that **Ct** polyamides exhibited specificity for T > A at the cap position with excellent polyamide affinities. We hoped that a hybrid dimer would impart excellent specificity for the TT sequence while maintaining a biologically relevant affinity – something that compounds containing multiple **Hp** were incapable of doing. Oligomer **30** bound its designed match site 5'-**TT**ACA-3' with a sub-nanomolar affinity of $K_a = 2.4 \times 10^9 \, M^{-1}$ and with an increased

Table 7.2. Affinities of Z/Py Ring Pairs Proximal to a Hydroxybenzimidazole Bicycle $K_a \, [M^{-1}]^{a,b}$

	5'-ATAC-3'	5'-TTAC-3'	5'-GTAC-3'	5'-CTAC-3'
Im/Py	$9.7\,(\pm0.7)\text{x}\,10^7$	$4.5\,(\pm0.6)\text{x}\,10^8$	$1.7\,(\pm0.4)\text{x}\,10^8$	$\leq 1.0 \times 10^7$
Tn/Py	$5.8\,(\pm0.8)\text{x}\,10^8$	$7.8\,(\pm0.5)\text{x}\,10^8$	$7.6\,(\pm0.5)\text{x}\,10^8$	$\leq 1.0 \times 10^7$
Ct/Py	$2.1\,(\pm0.3)\text{x}\,10^8$	$2.4\,(\pm0.2)\text{x}\,10^9$	$2.6\,(\pm0.4)\text{x}\,10^8$	$\leq 1.0 \times 10^7$
Is/Py	$1.1\,(\pm0.2)\text{x}\,10^9$	$1.0\,(\pm0.5)\text{x}\,10^9$	$2.0\,(\pm0.3)\text{x}\,10^9$	$6.6\,(\pm0.6)\text{x}\,10^7$
No/Py	$8.6\,(\pm0.3)\text{x}\,10^8$	$9.5\,(\pm0.3)\text{x}\,10^8$	$6.8\,(\pm0.4)\text{x}\,10^9$	$2.7\,(\pm0.5)\text{x}\,10^8$

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM MgCl$_2$, and 5 mM CaCl$_2$ at pH 7.0.

specificity of the chlorothiophene ring for T•A over A•T from ~4-fold in a **Ct**-Py system to 10-fold on the **Ct-Hz** system. This seemingly synergistic effect is attributed to the fact that both the exocyclic chlorine and hydroxyl groups prefer to lie across the less bulky thymine base in an A•T base pair and that the –OH of the **Hz** ring is able to form an energetically favorable hydrogen bond with the O2 carbonyl of thymine.[13] Combined,

these attributes makes this dimer the preferred system for targeting consecutive thymine residues.

We next looked to evaluate how the G specific **Im** ring would perform once
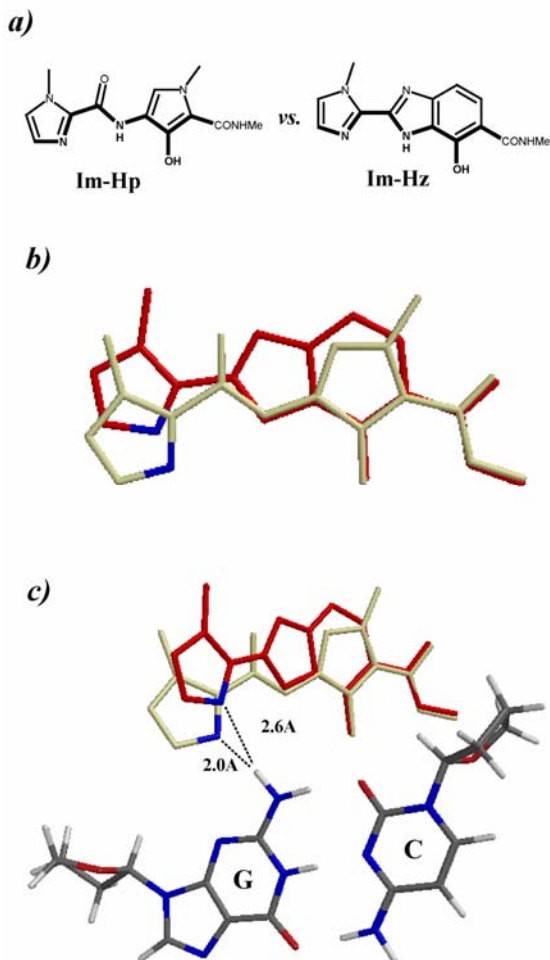

Figure 7.9. Comparative geometry of imidazole-hydroxypyrrole (**Im**-**Hp**) and imidazole-hydroxybenzimidazole (**Im**-**Hz**) dimeric units. (a) Structure of **Im**-**Hp** and **Im**-**Hz** with DNA groove contacts shown in bold. (b) Overlay shows **Im**-**Hp** in tan, **Im**-**Hz** in red, and the imidazole nitrogen in blue. Hydrogen not shown. (c) Overlay of **Im**-**Hp**:DNA and **Im**-**Hz**:DNA complexes showing the difference in distance of the imidazole nitrogen to the exocyclic amine of guanine. Coordinates for the **Im**-**Hp**:DNA complex taken from previously reported crystal structure data.[2]

incorporated in the **Im-Hz** dimer cap (polyamide **28**). We had hoped that the hairpin would prefer its match sequence of 5'-T**GT**ACA-3' and were surprised when it failed to demonstrate any preference for its designed site in addition to displaying a significantly decreased average affinity of ~10^8 M^{-1}. Since the specificity of the **Im** ring arises from its ability to accept a hydrogen bond from the exocyclic amine of guanine, we hypothesize that the less curved **Hz** ring could be responsible for the apparent loss of **Im** specificity. To test our hypothesis we took to modeling both the Im-H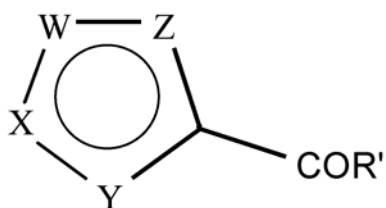p dimer and the **Im-Hz** dimer cap and compared the resulting differences in the **Im** position (Figure 7.9). As expected, linking the Im residue to the **Hz** 6-5 fused ring system

via a linear carbon-carbon bond results in a dimer that effectively pulls the terminal **Im** residue away from the minor groove floor and, in turn, away from the exocyclic amine of guanine (Figure 7.9b, red structure). Figure 7.9c illustrates how this structural change could affect the cap's ability to hydrogen bond, and thus, specifically bind across from guanine.

The shortcomings of the **Im-Hz** dimer prompted a search for a ring system that was capable of specifying for G > C within the **Xx-Hz** context. The isoxazole (**Is**) and oxazole (**No**) caps (Figure 7.1) were considered because of their structural resemblance to **Im** - all three rings present a nitrogen atom capable of hydrogen bonding to the minor groove. The binding profiles of both rings were accessed within the standard 8-ring hairpin design and it was found that **No** exhibited a 5.5-fold preference for G > C, while **Is** showed no appreciable specificity for any of the four base pairs (Figure 7.7). After evaluating the electron densities on the atom presented to the minor groove of several end-caps (Figure 7.10) it was found that of the nitrogen presenting rings, **No** presented the most electron-rich nitrogen to the minor groove and would be the most likely candidate to successfully hydrogen bond. Interestingly, **Is**, which was found to be degenerate at the cap position, was calculated to have even less electron density that **Im** thus backing the thermodynamics results. Further substantiating our claims was the

Ring	W	X	Y	Z	Charge on Y (e)
Im	C-H	C-H	N	N-Me	-0.54
Tn	C-H	C-H	S	C-Me	-0.09
Mt	C-H	C-H	**C-OMe**	S	+0.10
Ct	C-H	C-H	**C-Cl**	S	-0.10
Is	C-Me	O	N	C-H	-0.42
No	C-H	C-H	N	C-Me	-0.63

Figure 7.10. Structure of cap heterocycles and partial charge presented to the minor groove of the DNA by the recognition edge of each ring.
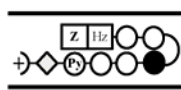
inability of the **Is-Hz** dimer to selectively target 5'-**TGT**ACA-3'. To our delight, when

the **No-Hz** dimer was incorporated into polyamide **32**, it was found to be specific for its

designed sequence of 5'-**TGT**ACA-3' with a 25-fold preference for G > C and an affinity

of $K_a = 6.8 \times 10^9$ M^{-1} at its match site (Figure 7.8). The **No-Hz** dimer presents the same

functionality to the minor groove as the **Im-Hz** dimer, but with an increased negative

charge on the guanine selective endocyclic nitrogen (Figure 7.10). This increased charge

may enhance its ability to hydrogen bond to the exocyclic amine of guanine, explaining

its ability to target a G-T site.

Conclusion.

Eight-ring hairpin oligomers containing the **No-Hz** and **Ct-Hz** dimer caps are able

to target G-T and T-T sequences at the oligomeric N-terminus with good specificity and

affinity and represent one of our most effective recognition systems (Table 7.3). The **No-Hz** and **Ct-Hz** dimer caps denote our first successful attempts to shift the focus of our recognition element designs away from single base pair interactions and towards oligomer/DNA sequence recognition.

Table 7.3. Specificities of Z/Py Ring Pairs Proximal to a Hydroxybenzimidazole Bicycle K_a $[M^{-1}]^{a,b}$

Z Hz / P	5'-ATAC-3'	5'-TTAC-3'	5'-GTAC-3'	5'-CTAC-3'
Im/Py	–	–	–	–
Tn/Py	–	–	–	–
Ct/Py	–	+	–	–
Is/Py	–	–	–	–
No/Py	–	–	+	–

recognition. We hope to apply this more comprehensive philosophy to future oligomer

designs with a focus on targeting sequences that have been problematic in the past.

Acknowledgments.

Experimental.

General. N,N-dimethylformamide (DMF), N,N-diisopropylethylamine (DIEA), thiophenol (PhSH), N,N-diethylamine, N,N-dimethylaminopropylamine (Dp), triethylamine (TEA), methyl-alpha-isocyanoacetate, and diazobicycloundecane (DBU) were purchased from Aldrich. Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam-resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), N,N-dimethylaminopyridine (DMAP), and Boc-β-alanine were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All other solvents were reagent grade from EM. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [γ-^{32}P]triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) and DNaseI (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA

polymerase was from Perkin-Elmer and used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d6* or CDCl$_3$ with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the California Institute of Technology. Matrix-assisted, laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was conducted at the Protein and Peptide Microanalytical Facility at the California Institute of Technology.

Heterocycle Synthesis. The synthesis of compounds **Im**-OH (**1**), **Tn**-OH (**2**), **Ct**-OH (**3**), **8**, **Im**-**Hz**(OMe)-OMe (**13**), **Im**-**Hz**(OMe)-OH) (**18**) and **33** (aryl diamine) have previously been reported (Figure 3).[15, 17] **Is**-OH (**4**) is commercially available.

5-methylisoxazole-3-carboxylic acid (Is-OH) **4***:* Physical data included for reference. TLC (3:2 EtOAc/Hex +10% AcOH) R_f 0.4; ^{1}H NMR (DMSO-d_6) 6.54 (s, 1H), 2.45 (s, 3H); [13]C (DMSO-d_6) 171.74, 161.21, 157.17, 102.47, 11.88; EI-MS m/e 127.027 (M[+] calcd. for 127.027 C$_5$H$_5$NO$_3$).

Methyl 5-methyl-1,3-oxazole-4-carboxylate (No-OMe) **6**: To a mixture of methyl-

alpha-isocyanoacetate (3 g, 30.2 mmol) and DBU (4.5 g, 30.2 mmol) in dry THF (40 mL)

at 10 ^{o}C, was added acetic anhydride (3.06 g, 3.0 mmol) in dry THF (10 mL) dropwise.

The reaction was allowed to warm to room temperature and stirred for 10 h. The solvent

was removed by rotoevaporation and water (100 mL) was added.   The mixture was

extracted with EtOAc (2 x 100 mL). The organic layer was collected and dried over

sodium sulfate. Removal of organics by rotoevaporation provided **6** as a crude amber oil.

The oil was subjected to column chromatography using (3:2 Hex/EtOAc) to provide **6**

(2.73 g, 64% Yield) as a crystalline white solid. TLC (3:2 Hex/EtOAc) R_f 0.5; ^{1}H NMR

(DMSO-d_6) 8.32 (s, 1H), 3.77 (s, 3H), 2.55 (s, 3H); ^{13}C (DMSO-d_6) 161.9, 156.1, 150.3,

126.2, 51.5, 11.5; EI-MS m/e 141.043 (M$^+$ calcd. for 141.043 $C_6H_7NO_3$).

5-Methyl-1,3-oxazole-4-carboxylic acid (No-OH) **7**:   A mixture of **6** (1 g, 7.08

mMol), 1N NaOH (10 mL) and MeOH (5 mL) was stirred at 40 ^{o}C for 4 h. The MeOH

was removed by rotoevaporation and the pH adjusted to pH = 2 with 1N HCl.   The

precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate

and removed by rotoevaporation to provide **7** (738 mg, 82% Yield) as a fibrous white

solid. TLC (3:2 EtOAc/Hex +10% AcOH) R_f 0.4; ^{1}H NMR (DMSO-d_6) 8.27 (s, 1H),

2.45 (s, 3H); ^{13}C (DMSO-d_6) 163.0, 155.6, 150.1, 145.6, 11.6; EI-MS m/e 127.027 (M$^+$

calcd. for 127.027 $C_5H_5NO_3$).

Tn-Hz(OMe)OMe **14**: To a solution of **2** (500 mg, 3.52 mmol) in DMF (6 mL)

was added DIEA (.676 mL, 3.87 mmol) and HBTU (1.26 g, 3.34 mmol). The mixture

was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **33**

(690 mg, 3.52 mmol). The reaction was then heated to 35 ^{o}C and stirred for an additional

24 h. The reaction was allowed to cool to room temperature and then poured into a Falcon tube containing cold water (40 mL) upon which time a cloudy precipitate formed. The Falcon tube was centrifuged at 14000 rpm for 10 min, the mother liquor decanted, and the precipitated solid dried under high vacuum. After drying, the crude solid (**9**) was dissolved in acetic acid (5 mL) and heated to 90 $^{\circ}$C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (3:2 EtOAc/Hex) to provide **14** as an off-white solid (447 mg, 42% Yield). TLC (3:2 EtOAc/Hex) R_f 0.6; ^{1}H NMR (DMSO-d_6) 7.64 (d, 1H, J = 4.8 Hz), 7.54 (d, 1H, J = 8.1 Hz), 7.21 (d, 2H, J = 8.1 Hz), 7.08 (d, 1H, J = 4.8 Hz) 4.32 (s, 3H), 3.78 (s, 3H), 2.59 (s, 3H); EI-MS m/e 302.072 (M$^+$ calcd. for 302.072 $C_{15}H_{14}N_2O_3S$).

Ct-Hz(OMe)OMe **15**: To a solution of **3** (294 mg, 1.80 mmol) in DMF (5 mL) was added DIEA (.348 mL, 1.98 mmol) and HBTU (650 mg, 1.72 mmol). The mixture was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **33** (300 mg, 1.80 mmol). The reaction was then heated to 35 $^{\circ}$C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a Falcon tube containing cold water (40 mL) upon which time a cloudy precipitate formed. The Falcon tube was centrifuged at 14000 rpm for 10 min, the mother liquor decanted, and the precipitated solid dried under high vacuum. After drying, the crude solid (**10**) was dissolved in acetic acid (5 mL) and heated to 90 $^{\circ}$C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (3:2 EtOAc/Hex) to provide **15** as an off-white solid (647 mg, 57% Yield). TLC (3:2 EtOAc/Hex) R_f 0.5; ^{1}H NMR (DMSO-d_6) 7.89 (d,

1H, J = 5.1 Hz), 7.57 (d, 1H, J = 8.4 Hz), 7.25 (m, 2H, J = 5.4), 4.28 (s, 3H), 3.79 (s, 3H); EI-MS m/e 322.017 (M$^+$ calcd. for 322.017 $C_{14}H_{11}N_2O_3SCl$).

Methyl   7-methoxy-2-(5-methylisoxazole-3-yl)benzimidazole-6-carboxylate   (Is-Hz(OMe)OMe) **16**: To a solution of **4** (0.3 g, 2.36 mmol) in DMF (4 mL) was added DIEA (915 mg, 1.23 mL, 7.08 mmol) and HBTU (895 mg, 2.36 mmol). The mixture was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **33** (463 mg, 2.36 mmol). The reaction was then heated to 35 ^{o}C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a Falcon tube containing cold water (40 mL) upon which time a cloudy precipitate formed. The Falcon tube was centrifuged at 14000 rpm for 10 min, the mother liquor decanted, and the precipitated solid dried under high vacuum. After drying, the crude solid (**11**) was dissolved in acetic acid (5 mL) and heated to 90 ^{o}C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (3:2 EtOAc/Hex) to provide **16** as a thin film. Addition of hexanes to the film, followed by rotoevaporation and drying under high vacuum provided **16** as a white solid (305 mg, 45% Yield). TLC (3:2 EtOAc/Hex) R_f 0.5; ^{1}H NMR (DMSO-d_6) 7.60 (d, 1H, J = 8.4 Hz), 7.21 (d, 1H, J = 8.4 Hz), 6.86 (s, 1H), 4.31 (s, 3H), 3.79 (s, 3H) 2.51 (s, 3H); ^{13}C (DMSO-d_6) 171.3, 168.0, 154.9, 151.7, 142.2, 141.0, 135.2, 127.1, 115.8, 105.8, 101.4, 61.2, 52.1, 11.8; EI-MS m/e 287.091 (M$^+$ calcd. for 287.090 $C_{14}H_{13}N_3O_4$).

Methyl   7-methoxy-2-(5-methyl(1,3-oxazole-4-yl))benzimidazole-6-carboxylate (NoHz(OMe)OMe) **17**: To a solution of **7** (0.3 g, 2.36 mmol) in DMF (4 mL) was added DIEA (915 mg, 1.23 mL, 7.08 mmol) and HBTU (895 mg, 2.36 mmol). The mixture was

stirred at room temperature for 1 h, followed by the addition of the aryl diamine **33** (463 mg, 2.36 mmol). The reaction was then heated to 35 $^{\circ}$C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a separatory funnel containing water (200 mL). The water was then extracted with EtOAc (2 x 100 mL). The organic layer was dried over sodium sulfate and removed by rotoevaporation to provide **12** as a crude residue. **12** was then dissolved in acetic acid (5 mL) and heated to 90 $^{\circ}$C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (4:1 EtOAc/Hex) to provide **17** as a thin film. Addition of hexanes to the film, followed by rotoevaporation and drying under high vacuum provided **17** as a white solid (379 mg, 56% Yield). TLC (4:1 EtOAc/Hex) R_f 0.75; ^{1}H NMR (DMSO-d_6) 8.49 (s, 1H), 7.51 (d, 1H, J = 8.7 Hz), 7.16 (d, 1H, J = 8.7 Hz), 4.34 (s, 3H), 3.77 (s, 3H) 2.78 (s, 3H); ^{13}C (DMSO-d_6) 166.7, 151.0, 149.3, 145.5, 139.0, 135.3, 125.6, 124.8, 114.2, 105.3, 60.8, 51.6, 11.3; EI-MS m/e 287.091 (M^{+} calcd. for 287.090 $C_{14}H_{13}N_3O_4$).

(Tn-Hz(OMe)OH) **19**: A mixture of **14** (500 mg, 0.97 mMol), 1N NaOH (8 mL) and MeOH (8 mL) was stirred at 40 $^{\circ}$C for 6 h. The MeOH was removed by rotoevaporation and the aqueous layer washed with EtOAc (3 x 10 mL). The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **19** (338 mg, 71% Yield) as an off-white solid. TLC (3:2 EtOAc/Hex + 10% AcOH) R_f 0.6; ^{1}H NMR (DMSO-d_6) 7.65 (d, 1H, J = 4.8 Hz), 7.55 (d, 1H, J = 8.1 Hz), 7.21 (d, 2H, J = 8.1), 7.08 (d, 1H, J = 4.8 Hz) 4.28 (s, 3H), 2.59 (s, 3H); EI-MS m/e 288.056 (M^{+} calcd. for 288.056 $C_{14}H_{12}N_2O_3S$).

(Ct-Hz(OMe)OH) **20**: A mixture of **15** (500 mg, 1.55 mMol), 1N NaOH (8 mL) and MeOH (8 mL) was stirred at 40 ^{o}C for 5 h. The MeOH was removed by rotoevaporation and the aqueous layer washed with EtOAc (3 x 10 mL). The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **20** (330 mg, 69% Yield) as a light-yellow solid. TLC (3:2 EtOAc/Hex + 10% AcOH) R_f 0.5; ^{1}H NMR (DMSO-d_6) 7.88 (d, 1H, J = 5.1 Hz), 7.59 (d, 1H, J = 8.4 Hz), 7.29 (d, 1H, J = 8.4), 7.25 (d, 1H, J = 5.1 Hz), 4.25 (s, 3H); EI-MS m/e 308.003 (M$^+$ calcd. for 308.002 C$_{13}$H$_9$N$_2$O$_3$SCl).

(Is-Hz(OMe)OH) **21**: A mixture of **16** (280 mg, 0.97 mMol), 1N NaOH (2 mL) and MeOH (1 mL) was stirred at 40 ^{o}C for 4 h. The MeOH was removed by rotoevaporation and the pH adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **21** (226 mg, 85% Yield) as a white solid. TLC (3:2 EtOAc/Hex +10% AcOH) R_f 0.4; ^{1}H NMR (DMSO-d_6) 7.62 (d, 1H, J = 8.7 Hz), 7.21 (d, 1H, J = 8.7 Hz), 6.87 (s, 1H), 4.30 (s, 3H), 2.51 (s, 3H); ^{13}C (DMSO-d_6) 171.2, 167.5, 155.3, 151.7, 142.0, 139.3, 135.3, 126.5, 115.7, 105.7, 101.0, 61.2, 11.8; EI-MS m/e 273.075 (M$^+$ calcd. for 273.075 C$_{13}$H$_{11}$N$_3$O$_4$).

7-methoxy-2-(5-methyl(1,3-oxazole-4-yl))benzimidazole-6-carboxylic   acid (NoHz(OMe)OH) **22:** A mixture of **17** (200 mg, 0.69 mMol), 1N NaOH (2 mL) and MeOH (1 mL) was stirred at 40 ^{o}C for 4 h. The MeOH was removed by rotoevaporation and the pH adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **22** (156 mg, 82% Yield) as a white solid. TLC (3:2 EtOAc/Hex +10% AcOH) R_f

0.35; ^{1}H NMR (DMSO-d_6) 8.50 (s, 1H), 7.54 (d, 1H, J = 8.4 Hz), 7.18 (d, 1H, J = 8.4 Hz), 4.29 (s, 3H), 2.78 (s, 3H); ^{13}C (DMSO-d_6) 168.1, 151.5, 149.8, 138.8, 136.1, 126.2, 125.6, 61.5, 11.8; EI-MS m/e 273.075 (M$^+$ calcd. for 273.075 $C_{13}H_{11}N_3O_4$).

Oligomer Synthesis: Oligomers were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid phase protocols.[18] The synthesis of batch resin **BR1** (H$_2$N-Py-Py-γ-Im-Py-Py-Py-β-Pam), **BR2** (H$_2$N-Py-Py-Py-γ-Im-Py-Py-Py-β-Pam) and polyamides **23**-**25** has previously been reported.[15] (Figure 4).

Is-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**26**):   (**Is**-OH) (11.2 mg, 88.5 µmol) was incorporated by activation with HBTU (32 mg, 84 µmol), DIEA (23 mg, 31 µl, 177 µmol) and DMF (250 µl).   The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing base resin **BR2** H$_2$N-Py-Py-Py-γ-Im-Py-Py-Py-β-Pam. Coupling was allowed to proceed for 12 h at 37 ^{o}C. The resin was then washed with DCM followed by the addition of Dp (500 µL). The mixture was allowed to stand for 2 h at 85 ^{o}C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC to provide **Is**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**26**) (2.8 mg, 7.7 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1223.56 (M+H calcd. for 1223.56 $C_{58}H_{71}N_{20}O_{11}$).

No-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**27**): (**No**-OH) was incorporated as described above for **Is**-OH (oligomer **26**) to provide **No**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**26**) (3.1

mg, 8.5 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1223.58 (M+H calcd. for 1223.56 $C_{58}H_{71}N_{20}O_{11}$).

Im-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**28**): (**Im**-**HzOMe**-OH) (25 mg, 88.5 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μl, 177 μmol) and DMF (250 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing base resin **BR1** H_2N-Py-Py-γ-Im-Py-Py-Py-β-Pam. Coupling was allowed to proceed for 12 h at room temperature. The resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide **Im**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**28**) (0.9 mg, 2.4 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1232.55 (M+H calcd for 1232.56 $C_{59}H_{70}N_{21}O_{10}$).

Tn-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**29**): (**Tn**-**HzOMe**-OH) was incorporated as described above for **Im**-**Hz**-OH (oligomer **28**) to provide **Tn**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**29**) (1.1 mg, 3.0 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1248.55 (M+H calcd. for 1248.53 $C_{60}H_{70}N_{19}O_{10}S$).

Ct-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**30**): (**Ct**-**HzOMe**-OH) was incorporated as described above for **Im**-**Hz**-OH (oligomer **28**) to provide **Ct**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**30**) (1.1 mg, 2.9 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1269.47 (M+H calcd. for 1269.47 $C_{59}H_{67}ClN_{19}O_{10}S$).

Is-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**31**): (**Is**-**HzOMe**-OH) was incorporated as described above for **Im**-**Hz**-OH (oligomer **28**) to provide **Is**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-

Dp (**33**) (0.8 mg, 2.2 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1233.55 (M+H calcd. for 1233.55 $C_{59}H_{69}N_{20}O_{11}$).

No-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**32**): (**No**-**HzOMe**-OH) was incorporated as described above for **Im**-**Hz**-OH (oligomer **28**) to provide **No**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**32**) (1.5 mg, 4.1 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1233.54 (M+H calcd for 1233.55 $C_{59}H_{69}N_{20}O_{11}$).

Deprotection of the *O*-Methyl-Protected Polyamides. O-Methyl protected polyamides were cleaved from resin, purified, deprotected and subjected to further purification using the following general procedure: Upon completion of solid-phase synthesis, Dp (500 μL) was added to the synthesis vessel containing the resin (50 mg). The mixture was allowed to stand for 2 h at 85 °C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC and lyophilized to provide polyamides containing the *O*-methyl protected hydroxybenzimidazole unit (-**HzOMe-**) as a dry solid. The polyamides were then dissolved in DMF (200 μl) and added to a suspension of sodium hydride (40 mg, 60% oil dispersion) and thiophenol (200 μl) in DMF (400 μl) that was pre-heated for 5 min at 85 °C. The mixture was heated for 2 h at 85 °C. The mixture was then cooled to 0 °C and 20% TFA (7.0 mL) was added. The aqueous layer was washed three times with diethyl ether (8 mL) and then diluted to a total volume of 9.5 mL using

0.1% TFA. The mixture was then purified by reverse-phase HPLC to give the deprotected **Hz**-containing oligomers **28**-**32**.

References:

[1] Trauger, J. W.; Dervan, P. B., *Drug-Nucleic Acid Interactions* **2001,** 340, 450-466.

[2] Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nature Structural Biology* **1998,** 5, (2), 104-109.

[3] Thrum, H., *Naturwissenschaften* **1959,** 46, (2), 87-87.

[4] Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X., *Journal of the American Chemical Society* **1951,** 73, (1), 341-343.

[5] 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, (5), 1376-1380.

[6] Pelton, J. G.; Wemmer, D. E., *Journal of the American Chemical Society* **1990,** 112, (4), 1393-1399.

[7] Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C., *Biochemistry* **1986,** 25, (23), 7408-7416.

[8] Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001,** 9, (9), 2215-2235.

[9] Helene, C., *Nature* **1998,** 391, (6666), 436-438.

[10] White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Nature* **1998,** 391, (6666), 468-471.

[11] Arcamone, F.; Nicolell.V; Penco, S.; Orezzi, P.; Pirelli, A., *Nature* **1964,** 203, (494), 1064.

[12] Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **1992,** 89, (16), 7586-7590.

[13] Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1999,** 121, (50), 11621-11629.

[14] Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B., *Helvetica Chimica Acta* **2002,** 85, (12), 4485-4517.

[15] Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2003,** 11, (20), 4333-4340.

[16] Briehn, C. A.; Weyermann, P.; Dervan, P. B., *Chemistry-a European Journal* **2003,** 9, (9), 2110-2122.

[17] Renneberg, D.; Dervan, P. B., *Journal of the American Chemical Society* **2003,** 125, (19), 5707-5716.

[18] Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1996,** 118, (26), 6141-6146.

[19] Urbach, A. R.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **2001,** 98, (8), 4343-4348.

[20] Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., *Journal of Molecular Biology* **2002,** 320, (1), 55-71.