

Chapter 6

Expanding the Repertoire of Heterocycle Ring Pairs for Programmable Minor Groove DNA Recognition

The text of this chapter was taken in part from a manuscript coauthored with Raymond M. Doss, Shane Foister and Professor Peter B. Dervan (Caltech)

(Marques, M. A.; Doss, R. M.; Foister, S. and Dervan, P. B. “*Expanding the Repertoire of Heterocycle Ring Pairs for Programmable Minor Groove DNA Recognition*” *Journal of the American Chemical Society* **2004**, 126, (33), 10339-10349.)

Abstract.

The discrimination of the four Watson-Crick base pairs by minor groove DNA binding polyamides have been attributed to the specificity of three 5-membered aromatic amino acid subunits, 1-methyl-1*H*-imidazole (**Im**), 1-methyl-1*H*-pyrrole (**Py**), and 3-hydroxy-1*H*-pyrrole (**Hp**) paired four different ways. The search for additional ring pairs that demonstrate DNA sequence specificity has led us to a new class of 6-5 fused bicycle rings as minor groove recognition elements. The affinities and specificities of the hydroxybenzimidazole/pyrrole (**H_z/Py**) and hydroxybenzimidazole/benzimidazole (**H_z/Bi**) pairs for each of the respective Watson-Crick base pairs within the sequence context 5'-TGG**X**CA-3' (**X** = A, T, G, C) were measured by quantitative DNase I footprinting titrations. The **H_z/Py** and **H_z/Bi** distinguish T•A from A•T. Hairpin polyamides containing multiple **H_z/Py** pairs were examined and shown to mimic the **Hp/Py** pair with regard to affinity and specificity. Therefore the **H_z/Py** pair may be considered a second-generation replacement for the **Hp/Py** pair.

Introduction.

Aberrant gene expression is the cause of many diseases and the ability to reprogram transcriptional pathways using cell-permeable small molecules may one day have an impact on human medicine.^{2, 3} DNA-binding polyamides, which are based on the architecture of the natural products netropsin and distamycin A,⁴⁻⁷ are capable of distinguishing all four Watson-Crick base pairs in the DNA minor groove.⁸⁻¹⁷ Sequence-specific recognition of the minor groove of DNA arises from the pairing of two different antiparallel 5-membered heterocyclic amino acids and the interplay of a variety of direct and indirect recognition elements.^{8, 10, 18} The direct readout, or information face, on the inside of the crescent shaped polyamide may be programmed by the incremental change of atoms on the corners of the ring pairs presented to the DNA minor groove floor. Stabilizing and destabilizing interactions with the different edges of the four Watson-Crick bases are modulated by shape complementarity and by specific hydrogen bonds.¹⁹⁻²³ For example, the imidazole ring **Im**, which presents a lone pair of electrons to the DNA minor groove, can accept a hydrogen bond from the exocyclic amine of guanine.²⁰ Additionally, the 3-hydroxypyrrole ring **Hp** projects an exocyclic OH group toward the minor groove floor that is sterically accommodated in the cleft of the T·A base pair, preferring to lie opposite T, not A. From x-ray structural analysis, it appears that **Hp** can form two hydrogen bonds with the O(2) of thymine.^{21, 24} Molecular recognition of DNA by polyamides is also affected by a series of critical ligand-DNA interactions that take place away from the polyamide recognition face. Polyamide geometry with respect to overall curvature, rise per residue, and contacts between the polyamide and the walls of the minor groove, as well as sequence-dependent energetic penalties that arise from distorting DNA from its low energy unbound conformation, are examples of how

sequence-dependent microstructure and flexibility of DNA may influence indirectly the affinity and specificity of polyamides.^{22, 23}

Previously, pairings using **Im**, **Py**, and **Hp** have been used to discriminate the Watson-Crick base pairs such that **Im/Py** is specific for G·C and **Hp/Py** for T·A.⁸⁻¹⁰ These pairing rules have proven useful for the recognition of hundreds of DNA sequences by these programmable oligomers. However, sequencedependent changes in the microenvironment of the DNA makes the targeting of certain sequences difficult, leading us to explore whether other novel heterocyclic recognition elements could be discovered for use in DNA groove recognition within the pairing paradigm.²⁵⁻²⁹

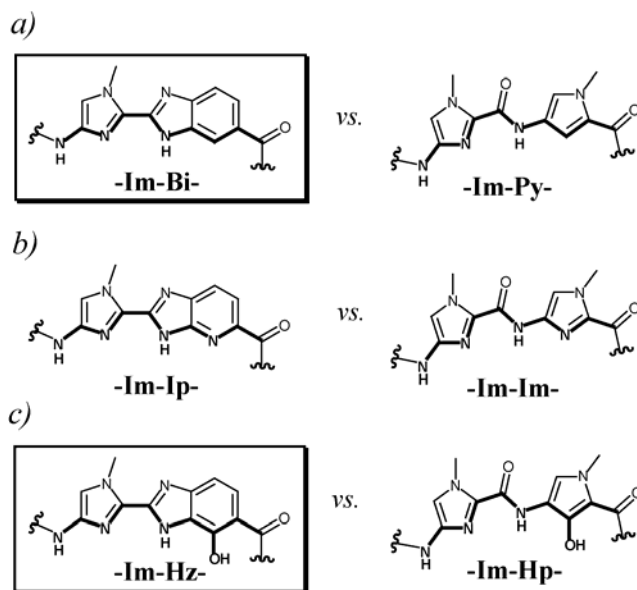


Figure 6.1. Structures of the (a) benzimidazole (Bi), (b) imidazolopyridine (Ip), and (c) hydroxybenzimidazole (Hz) building blocks in comparison with their respective 5-membered ring systems. Hydrogen bonding surfaces presented to the DNA minor groove are bolded.

Furthermore, replacement of the **Hp** ring system, which was a benchmark for the field with regard to its ability to distinguish T·A from A·T, was a priority due to the subsequent observation that hairpin oligomers containing **Hp** degrade over time in the presence of acid or trace free radicals.³⁰⁻³² We recently reported that the benzimidazole architecture

can be an effective platform for the development of novel modular recognition elements for the minor groove of DNA.³³⁻³⁵ The benzimidazole 6-5 bicyclic-ring structure, while

having different curvature from the classic 5-membered heterocyclic carboxamides, presents an “inside edge” with a similar readout shape to the DNA minor groove floor, effectively mimicking **Py**, **Im**, and **Hp** heterocycle-carboxamides (Figure 6.1).

Here we compare the hydroxybenzimidazole bicycle **Hz** with respect to the **Hp** amino acid at discriminating between the Watson-Crick bases within several different sequence contexts. Experiments were designed to elucidate the DNA recognition properties of the **Hz/Py** pair, and the ability of multiple **Hz/Py** pairs to distinguish

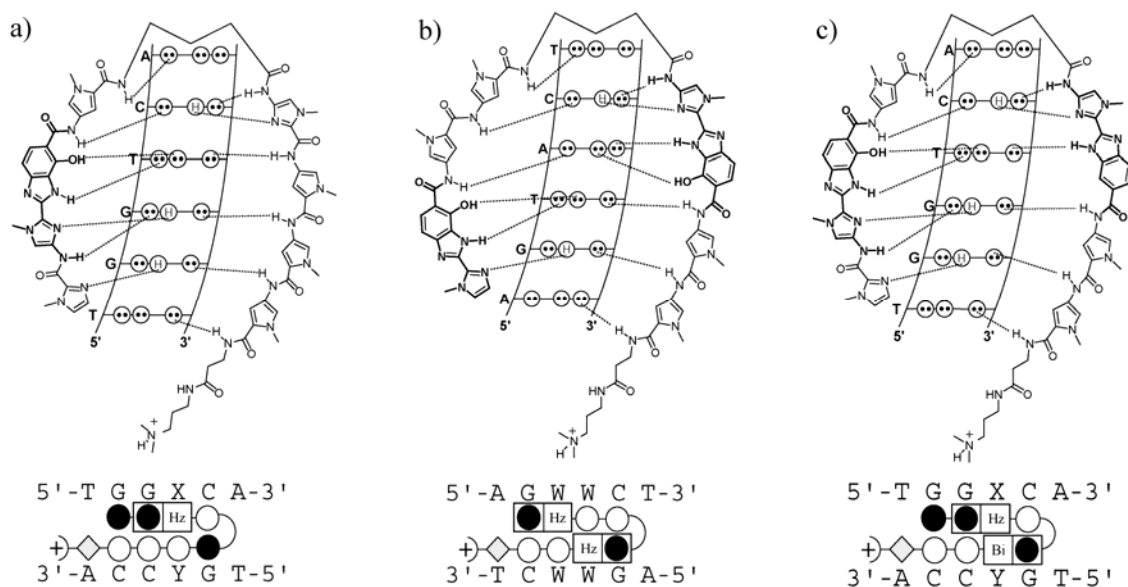


Figure 6.2. Hairpin polyamides and the minor groove contacts for their match sequences along with experimental schemes showing the DNA sequences that each polyamide will be tested against. (a) Hz/Py and Py/Hz pairs tested against the four Watson-Crick base pairs. (b) Multiple Hz/Py pairs tested against different core A,T sequences. (c) Hz/Bi and Bi/Hz pairs tested against the four Watson-Crick base pairs. Dark circle = imidazole; Light circle = pyrrole; Half circle = γ ; Diamond = β -Alanine; Half circle with a plus = dimethylaminopropylamide tail.

multiple A·T sequences. We also report that the 6-5/6-5 bicyclic pair **Hz/Bi** is effective for minor groove recognition (Figure 6.2). For the series of experiments, this required the synthesis of two new fused ring dimers **Bi-Im** and **Hz-Im** that would mimic the **Py-Im** and **Hp-Im** dimers. These were incorporated into the synthesis of hairpin oligomers **1–6** (Figure 6.3). DNaseI footprinting titrations were used to determine DNA binding

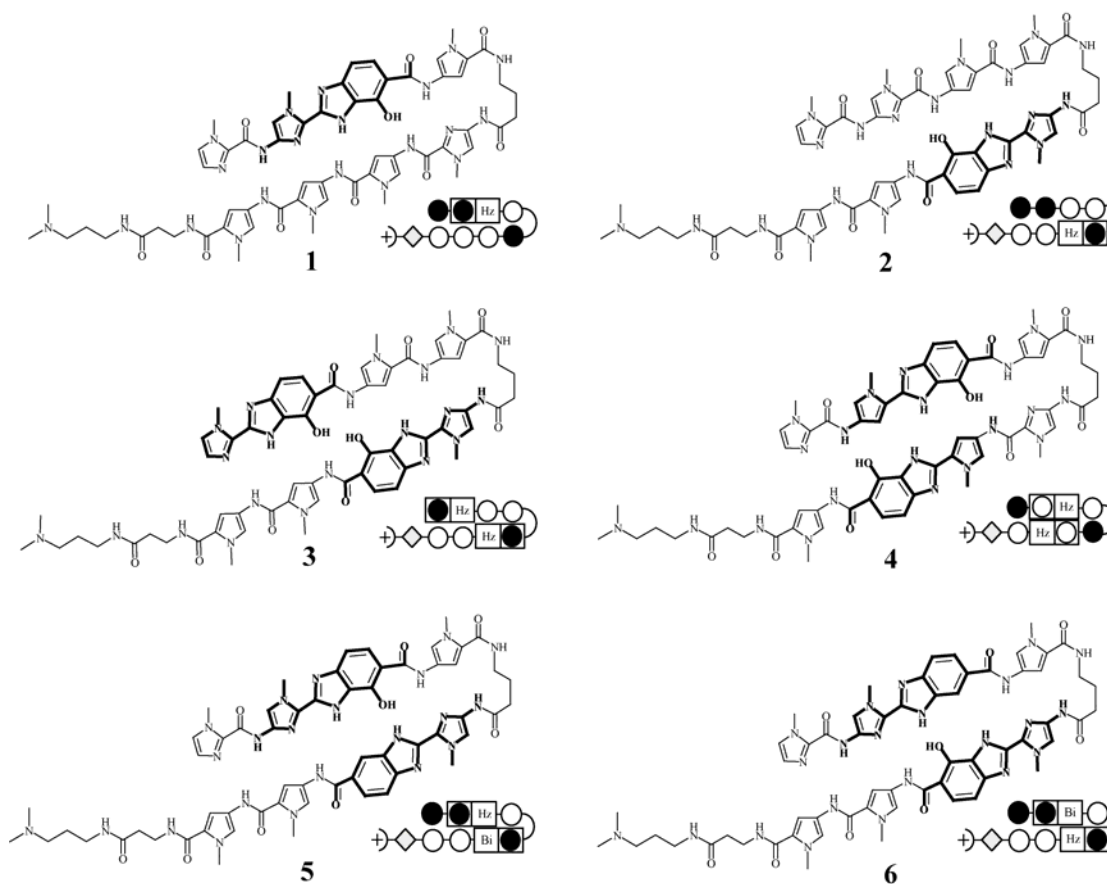


Figure 6.3 Structures of the polyamides containing Hz/Py, Hz/Bi, and multiple Hz/Py pairs shown along with their ball and stick representation. Shaded and non-shaded circles indicate imidazole and pyrrole respectively, whereas hydroxybenzimidazole and benzimidazole are indicated as Hz and Bi. The half circle represents γ -aminobutyric acid linker, while the diamond indicates β -Alanine. The half circle with a plus indicates the dimethylaminopropylamide tail.

affinities of hairpins **1–6**, which will reveal the energetic preferences of the **Hz/Py** and **Hz/Bi** pairs for the four Watson-Crick bases, as well as determine the fidelity of hairpin polyamides containing multiple **Hz/Py** pairs. Molecular modeling was used to further analyze the binding data.

Results.

Heterocycle Synthesis. Boc-protected heterocyclic amino acids **Py** and **Im** were synthesized as previously described.³⁶ Benzimidazole analogues Boc-Im-**Bi**-OH (**11**) and Boc-Py-**H**zOMe-OH (**12**) were also synthesized according to published protocols.³⁵ Compounds Im-**H**zOMe-OH (**16**) and Boc-Im-**H**zOMe-OH (**18**) were synthesized from diamine (**13**) (Figure 6.4).

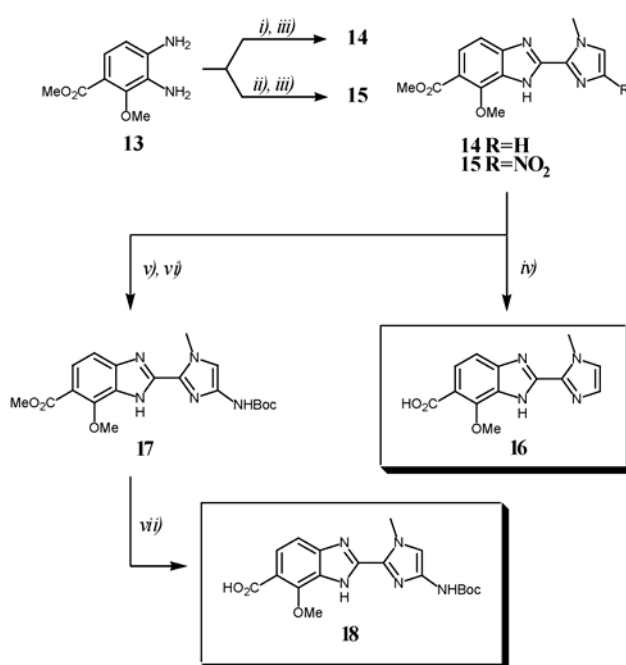


Figure 6.4. Synthesis of Im-**H**zOMe-OH (**16**) and Boc-Im-**H**zOMe-OH (**18**); (i) NO₂-Im-OH (**19**), HBTU, DIEA, DMF; (ii) Im-COCCl₃ (**9**), DIEA, EtOAc; (iii) AcOH; (iv) NaOH, MeOH; (v) H₂ Pd/C, DIEA, DMF; (vi) (Boc)₂O; (vii) NaOH, MeOH.

Im-H**zOMe-OH (**16**).** To a solution of diamine (**13**) in EtOAc was added Im-COCCl₃ (**9**) and DIEA. The mixture was heated at 35 °C and stirred for 3 h at which time a precipitate formed. The precipitate was filtered, washed, and dissolved in neat AcOH. The mixture was heated at 90 °C to provide the product Im-**H**zOMe-OH (**16**) cleanly. The formation

of the cyclocondensation product (**14**) can be followed easily by TLC and the appearance of a bright fluorescent spot under UV visualization. Im-**H**zOMe-OH (**16**) was then saponified using 1N NaOH in MeOH to provide Im-**H**zOMe-OH (**16**). The above synthetic route is attractive due to the absence of chromatography steps. Furthermore,

the trichloro ketone functionality of Im-COCCl₃ works well as an activated leaving group, providing condensation products at mild temperatures and neutral conditions.³⁷⁻³⁹

Boc-Im-HzOMe-OH (18). Diamine (**13**) was coupled to NO₂-Im-OH (**19**) using HBTU and DIEA in DMF. The mixture was stirred at ambient temperature for 24 h and then poured into water to produce a precipitate. The precipitate was collected, dissolved in neat AcOH, and heated at 90 °C to provide the cyclocondensation product NO₂-Im-HzOMe-OMe (**15**). Reduction of the imidazole nitro group was accomplished using H₂ Pd/C in a mixture of DMF and DIEA, followed by the addition of (Boc)₂O to provide the Boc-protected amino ester Boc-Im-HzOMe-OMe (**17**). **17** was saponified using a mixture of 1N NaOH and MeOH to provide the final product Boc-Im-HzOMe-OH (**18**). The synthesis of NO₂-Im-HzOMe-OMe (**15**) was attempted using the trichloro ketone protocol provided above for **14** but was unsuccessful. It appears the additional reactivity of the NO₂-Im-COCCl₃ species results in di-acylation of the diamine (**13**) and prohibits cyclocondensation.

Polyamide Synthesis. Polyamides **1-6** were synthesized in stepwise fashion on β-Pam resin following manual solid-phase methods.³⁶ The Boc-protected amino acids utilized for polyamide synthesis were Boc-Py-OBt (**7**), Boc-Im-OH (**8**), Im-COCCl₃ (**9**), Boc-γ-OH (**10**), Boc-Im-Bi-OH (**11**), Boc-Py-HzOMe-OH (**12**), Im-HzOMe-OH (**16**) and Boc-Im-HzOMe-OH (**18**) (Figure 6.5). Couplings were realized using pre-activated monomers (**7**) or HBTU activation in a DIEA and DMF mixture. Coupling times ran from 3-24 h at 25-40 °C. Deprotection of the growing polyamide was accomplished using 80% TFA/DCM. Polyamides were cleaved from the resin by treatment with dimethylaminopropylamine (Dp) neat at 80 °C for 2 h, and purified by preparatory reverse phase HPLC. Deprotection of the methoxy-protected polyamides was done using

a mix of thiophenoxide in DMF at 80 °C, to provide the free hydroxy derivatives after a second HPLC purification: Im-Im-**H**z-Py- γ -Im-**P**y-Py-Py- β -Dp (**1**), Im-Im-**P**y-Py- γ -Im-**H**z-Py-Py- β -Dp (**2**), Im-**H**z-**P**y-Py- γ -Im-**H**z-**P**y-Py- β -Dp (**3**), Im-**P**y-**H**z-Py- γ -Im-**P**y-**H**z-Py- β -Dp (**4**), Im-Im-**H**z-Py- γ -Im-**B**i-Py-Py- β -Dp (**5**) and Im-Im-**B**i-Py- γ -Im-**H**z-Py-Py- β -Dp (**6**).

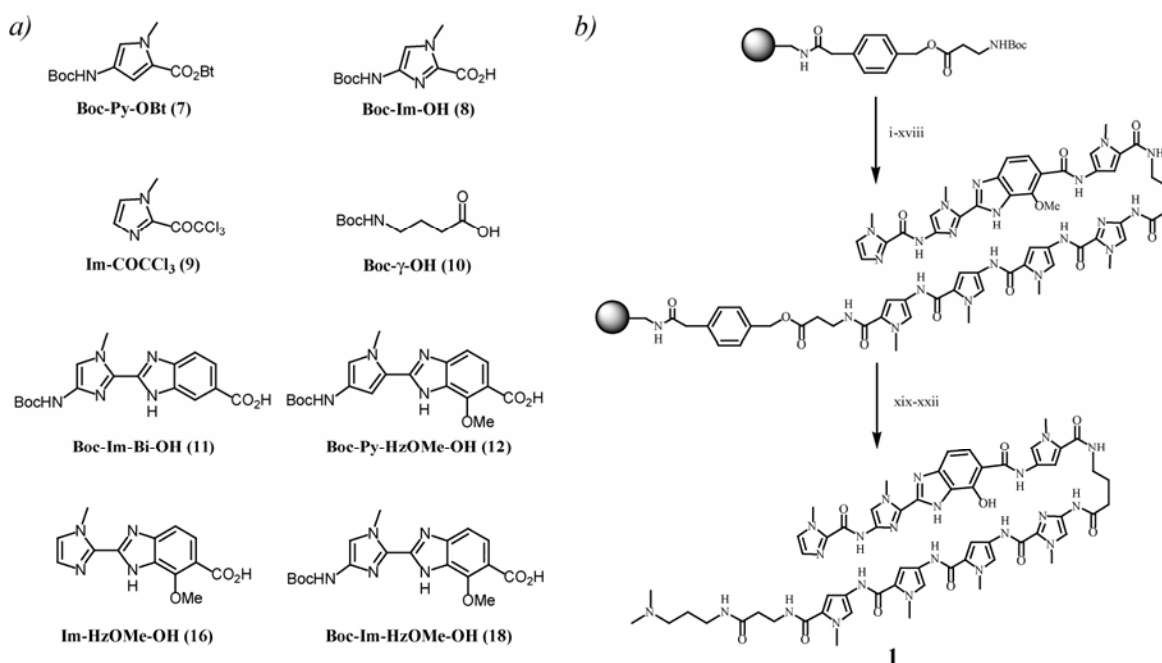


Figure 6.5. (a) Amino acid building blocks for polyamide synthesis. (b) Representative solid phase synthesis of polyamide 1. Reaction conditions: (i) 80% TFA/DCM; (ii) Boc-Py-OBt, DIEA, DMF; (iii) Ac₂O, DIEA, DMF; (iv) Repeat i-iii x2; (v) 80% TFA/DCM; (vi) Boc-Im-OH, HBTU, DIEA, DMF; (vii) Ac₂O, DIEA, DMF; (viii) 80% TFA/DCM; (ix) Boc- γ -OH, HBTU, DIEA, DMF; (x) Ac₂O, DIEA, DMF; (xi) 80% TFA/DCM; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) Ac₂O, DIEA, DMF; (xiv) 80% TFA/DCM; (xv) Boc-Im-HzOMe-OH, HBTU, DIEA, DMF; (xvi) Ac₂O, DIEA, DMF; (xvii) 80% TFA/DCM; (xviii) Im-COCCl₃, DIEA, DMF; (xix) dimethylaminopropylamine (Dp), 80 °C 2 h.; (xx) prep. HPLC; (xxi) thiophenol, NaH, DMF, 80 °C 2 h.; (xxii) prep. HPLC.

DNA Affinity and Sequence Specificity. Quantitative DNase-I-footprinting titrations were carried out for polyamides **1-6**. Polyamides **1**, **2**, **5**, and **6** were footprinted

on the 278-base-pair PCR product of plasmid pDHN1 (Figure 6.6a). Polyamides **3** and **4** were footprinted on the 285-base-pair PCR product of plasmid pDEH10 (Figure 6.6b).

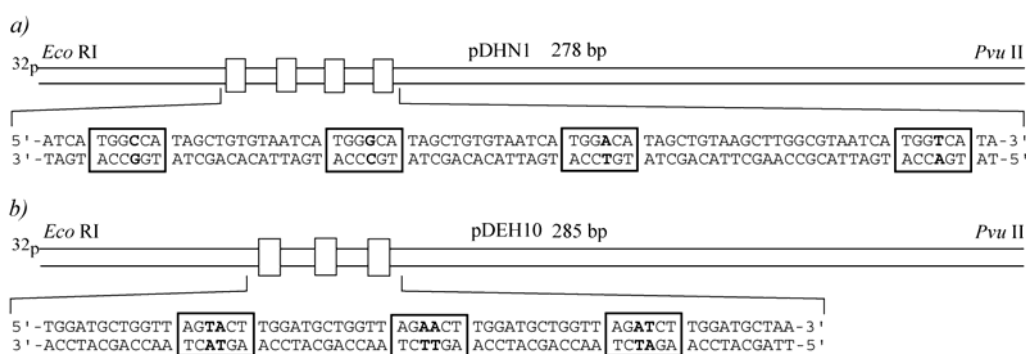


Figure 6.6. Illustration and complete sequence of the *EcoRI/PvuII* restriction fragment derived from plasmids (a) pDHN1 and (b) pDEH10. For pDHN1, the four designed 6-base-pair binding sites that were analyzed in quantitative DNaseI-footprinting titrations are shown with the variable Watson-Crick base pairs bolded and the binding site boxed. For pDEH10, the three designed 6-base-pair binding sites that were analyzed in quantitative DNaseI-footprinting titrations are shown with the variable Watson-Crick base pairs bolded and the binding site boxed.

For polyamides Im-Im-**H_z**-Py-γ-Im-**Py**-Py-Py-β-Dp (**1**) and Im-Im-**Py**-Py-γ-Im-**H_z**-Py-Py-β-Dp (**2**), the DNA-sequence specificity at a single ring-pairing position (bolded in the sequences listed above) was determined by varying a single DNA base pair within the parent-sequence context, 5'-TGG**X**CA-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (Figure 6.7). The variable base-pair position was installed opposite the novel **H_z/Py** and **Py/H_z** pairs, designed to target T·A and A·T respectively. Equilibrium association constants (K_a) for 8-ring polyamides containing **Py/Py**, **Hp/Py**, and **Py/Hp** pairs have been

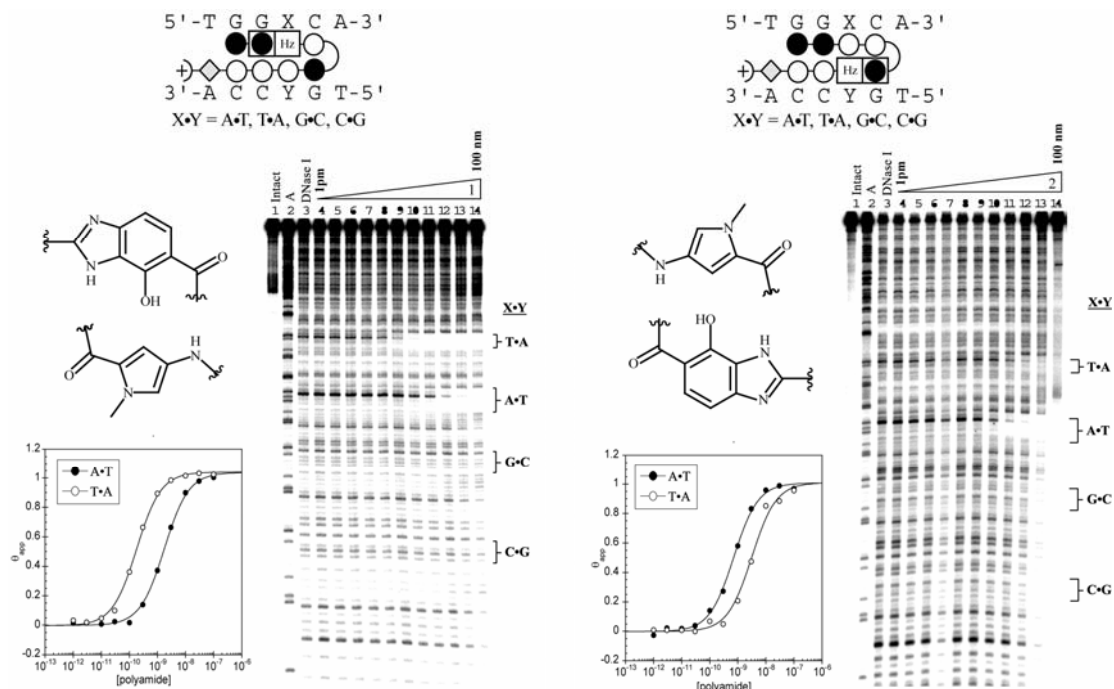







Figure 6.7. Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **1** and **2**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid DHN1: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the pairing of interest; and (bottom left) 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; Beta-alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; the hydroxybenzimidazole residue is indicated by a square containing Hz.

reported and are included for comparison with values presented here (Table 1).³¹

Polyamide **1** (**Hz/Py** pair) bound at high affinity and demonstrated a strong single-site specificity, preferring T•A over A•T by 10-fold, and A,T over G,C by more than 50-fold. In comparison to the **Hp/Py** pair, the **Hz/Py** exhibited a higher affinity, similar T vs. A specificity, and much greater A,T over G,C specificity. Polyamide **2** (**Py/Hz** pair) bound at high affinity and demonstrated a moderate single site specificity, preferring A•T over T•A by more than 4-fold, and A,T over G,C by more than 30-fold. In comparison to the **Py/Hp** pair, the **Py/Hz** pair within this sequence context shows

slightly lower T vs. A specificity, but improved A,T over G,C specificity. Both polyamides **1** and **2**, containing the **H_z/P_y** and **P_y/H_z** pairs, respectively, bound with

Table 6.1. *Hydroxypyrrrole & Hydroxybenzimidazole Hairpins: K_a [M^{-1}]^{a,b}*

Polyamide	A•T	T•A	G•C	C•G
 19	3.1 (±0.7) × 10 ⁹	4.7 (±0.4) × 10 ⁹	2.2 (±0.6) × 10 ⁸	2.5 (±0.9) × 10 ⁸
 20	8.1 (±1.9) × 10 ⁷	1.6 (±0.3) × 10 ⁹	5.5 (±1.5) × 10 ⁷	7.9 (±2.1) × 10 ⁷
 1	5.7 (±0.4) × 10 ⁸	5.5 (±0.2) × 10 ⁹	≤ 1.0 × 10 ⁷	≤ 1.0 × 10 ⁷
 21	1.1 (±0.2) × 10 ⁹	9.8 (±0.9) × 10 ⁷	2.5 (±0.3) × 10 ⁷	3.3 (±1.0) × 10 ⁷
 2	1.4 (±0.3) × 10 ⁹	3.2 (±0.6) × 10 ⁸	≤ 1.0 × 10 ⁷	≤ 1.0 × 10 ⁷

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂ at pH 7.0.

comparable affinity and single-site specificity in comparison to the degenerate **P_y/P_y** pair.

For polyamides Im-**H_z**-**P_y**-P_y-γ-Im-**H_z**-**P_y**-P_y-β-Dp (**3**) and Im-**P_y**-**H_z**-P_y-γ-Im-**P_y**-**H_z**-P_y-β-Dp (**4**) the ability of multiple **H_z/P_y** pairs to distinguish multiple A,T sequences was tested by varying two DNA base pairs across from the **H_z/P_y** and **P_y/H_z** pairs (bolded in the sequences listed above) within the parent-sequence context, 5'-AGWWCT-3', (W = A, T) and comparing the relative affinities of the resulting complexes (Figure 6.8). Equilibrium association constants (K_a) for 8-ring polyamides containing multiple **H_p/P_y** pairs have been reported and are included for comparison with values presented here (Table 6.2).⁴⁰ Polyamide **3** bound at moderate affinity and showed good specificity (greater than 14-fold) for its match sequence 5'-AGTACT-3' over 5'-AGAACT-3' and 5'-AGATCT-3'. Polyamide **4** bound all three sites at moderate affinity, demonstrating poor site selectivity. The recognition profiles of polyamides containing multiple **H_z/P_y** pairs (**3** and **4**) are similar to those reported for multiple **H_p/P_y** pairs. Polyamides **3** and **4** also demonstrated lowered affinities but superior

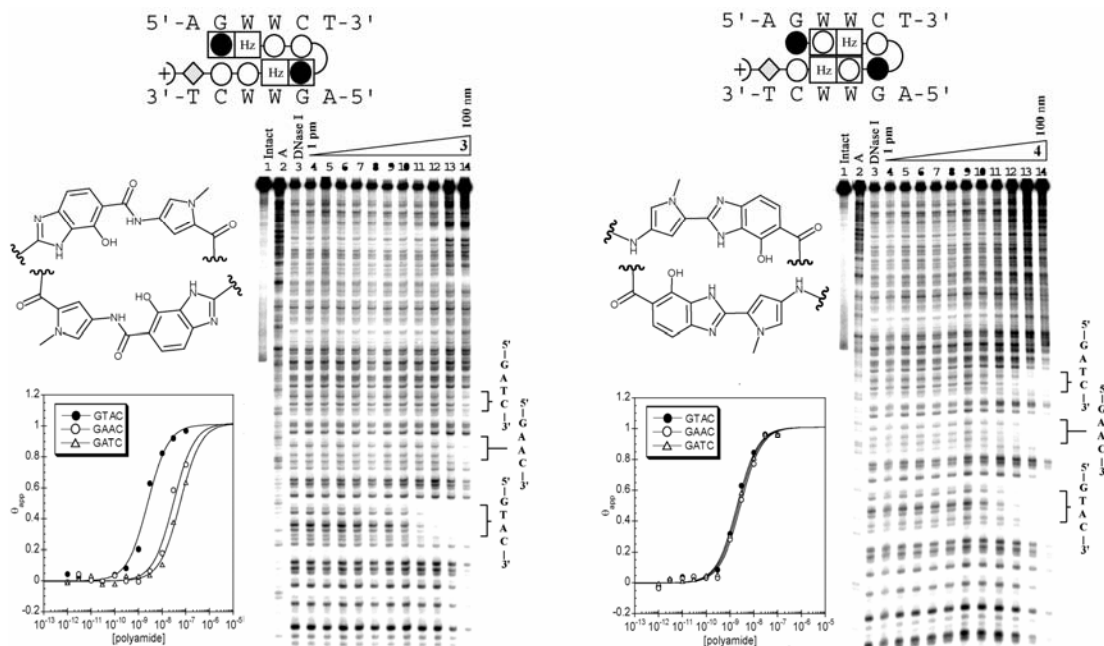







Figure 6.8. Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **3** and **4**, respectively, on the 285 bp, 5'-end-labelled PCR product of plasmid DEH10: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the pairing of interest; and (bottom left) 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; beta-alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; the hydroxybenzimidazole residue is indicated by a square containing Hz.

specificity in comparison to the comparator polyamide containing multiple symmetrical **Py/Py** pairs.

For polyamides Im-Im-**H**z-Py- γ -Im-**B**i-Py-Py- β -Dp (**5**) and Im-Im-**B**i-Py-Im-**H**z-Py-Py- β -Dp (**6**) the DNA-sequence specificity at a single ring-pairing position (bolded in the sequences listed above) was determined by varying a single DNA base pair within the parent-sequence context, 5'-TGG**X**CA-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (Figure 6.9 and Table 6.3). The variable base-pair position was installed opposite the novel **H**z/**B**i and **B**i/**H**z pairs, designed to target T•A and A•T, respectively. Polyamide **5**

(**H_z/B_i** pair) bound with a markedly high affinity, demonstrating a 13-fold selectivity for

Table 6.2. Multiple Hydroxypyrrrole & Hydroxybenzimidazole Ring Pairings: K_a [M^{-1}]^{a,b}

Polyamide	5'-aGTACT-3'	5'-aGAACT-3'	5'-aGATCt-3'
 22	$3.5 (\pm 0.7) \times 10^{10}$	$4.7 (\pm 0.7) \times 10^9$	$7.4 (\pm 1.5) \times 10^8$
 23	$7.0 (\pm 1.8) \times 10^8$	$\leq 1.0 \times 10^7$	$\leq 1.0 \times 10^7$
 3	$4.6 (\pm 0.8) \times 10^8$	$3.2 (\pm 0.4) \times 10^7$	$1.7 (\pm 0.5) \times 10^7$
 24	$1.0 (\pm 0.2) \times 10^8$	$2.6 (\pm 0.6) \times 10^7$	$3.3 (\pm 0.7) \times 10^7$
 4	$4.5 (\pm 0.7) \times 10^8$	$3.3 (\pm 0.7) \times 10^8$	$4.4 (\pm 0.9) \times 10^8$

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 MgCl₂, and 5 mM CaCl₂ at pH 7.0.

A,T over G,C and 2.5-fold preference for T·A over A·T. Polyamide **6** (**B_i/H_z** pair) also bound with high affinity, lowered A,T over G,C selectivity (4.5-fold), and a 2.5-fold preference for A·T over T·A. Polyamides containing **H_z/B_i** pairs bind with significantly higher affinity than those containing the **H_z/P_y** pairings, but with a mild loss in A,T selectivity.

Molecular Modeling. Modeling calculations were performed with the Spartan Essential software package.⁴¹ *Ab initio* calculations were done using a Hartree-Fock model and a 6-31G* polarization basis set. Four-ring subunits containing the sequence Im-Im-X-Py (**X** = **H_p**, **H_z**) were constructed to examine respective overall ligand geometry and curvature (Figure 6.10). Furthermore, dimeric subunits containing the sequence Im-X (**X** = **H_p**, **H_z**, **P_y**, and **B_i**) were constructed and evaluated using electron density, and isopotential plots (Figures 6.11 and 6.12).

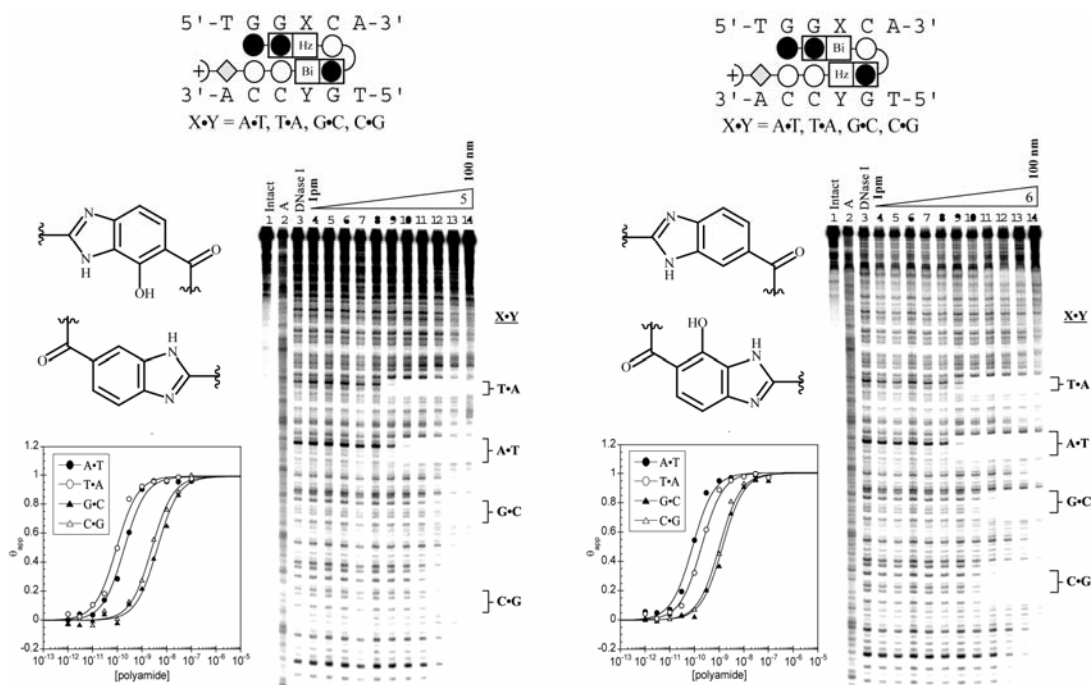


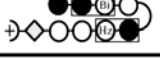


Figure 6.9. Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **5** and **6**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid DHN1: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the pairing of interest; and (bottom left) 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; beta-alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; the hydroxybenzimidazole residue is indicated by a square containing Hz; the benzimidazole residue is indicated by a square containing Bi.

Discussion. Previously, **Hp/Py** pairs have been the state of the art for distinguishing between T and A Watson-Crick bases, with **Hp/Py** coding for T•A and **Py/Hp** coding for A•T.⁹ The specificity imparted by **Hp** is attributed to the interaction of the exocyclic hydroxyl group with the asymmetric A,T base-pair cleft. More specifically, **Hp** functions through two modes: first, the hydroxyl group provides a steric bump that is better accommodated across from the thymine base, and second, the

hydroxyl proton may make a hydrogen bond with the O(2) of thymine. While **Hp** allows

Table 6.3. *Hydroxybenzimidazole/Benzimidazole Pairings: K_a [M^{-1}]^{ab}*

Polyamide	A•T	T•A	G•C	C•G
 19	$3.1 (\pm 0.7) \times 10^9$	$4.7 (\pm 0.4) \times 10^9$	$2.2 (\pm 0.6) \times 10^8$	$2.5 (\pm 0.9) \times 10^8$
 5	$4.1 (\pm 0.4) \times 10^9$	$1.0 (\pm 0.3) \times 10^{10}$	$2.4 (\pm 0.7) \times 10^8$	$3.2 (\pm 0.5) \times 10^8$
 6	$1.1 (\pm 0.4) \times 10^{10}$	$4.5 (\pm 0.4) \times 10^9$	$8.1 (\pm 0.8) \times 10^8$	$1.0 (\pm 0.7) \times 10^9$

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂ at pH 7.0.

for single base specificity, utilization of the hydroxypyrrole ring system results in polyamides with lower affinities, only moderate success at discriminating between multiple A,T sequences, and chemical instability over extended periods of time in solution. This motivated us to explore the molecular recognition capability of polyamides containing novel **H_z/Py** and **H_z/Bi** pairs.

Hydroxybenzimidazole/Pyrrole Pair (H_z/Py). Designed to be a hydroxypyrrole mimic, the **H_z/Py** pair places the same direct readout functionality to the floor of the DNA minor groove as the **Hp/Py** pair. Like the **Hp/Py** pair, the **H_z/Py** pair is capable of

Table 6.4. *Specificity of the H_z/Py and H_z/Bi Pairings^a*

Pair	A•T	T•A	G•C	C•G
H _z /Py	-	++	-	-
Py/H _z	++	-	-	-
H _z /Bi	-	+	-	-
Bi/H _z	+	-	-	-

a) Highly Favored (++), Favored (+), Disfavored (-)

effectively discriminating between A,T Watson-Crick base pairs such that **H_z/Py** codes for T•A and **Py/H_z** codes for A•T (Table 6.4).

The **H_z/Py** pair demonstrates an increase in binding affinity for its match sites and is comparable in specificity to the **Hp/Py** pairing. The **H_z/Py** pair also discriminates more effectively against G,C base pairs than **Hp/Py**. While the substituents presented to the floor of the DNA minor groove are virtually identical between **H_z** and **Hp**, there are significant differences in overall

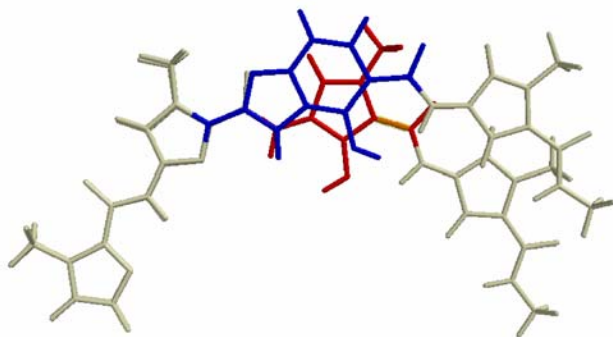


Figure 6.10. Schematic illustrating the respective curvatures of four-ring polyamide subunits Im-Im-Hp-Py and Im-Im-Hz-Py. Overall polyamide curvature effects how well the ligand can track the DNA minor groove. Hypercurvature negatively effects polyamide binding while curvature more complimentary to that of the DNA helix may allow for recognition of longer DNA sequences.

ligand geometry (Figure 6.10) and electronics (Figure 6.11). A higher degree of rigidity and preorganization of the fused hydroxybenzimidazole structure, coupled with a lower degree in curvature that may be more complementary to the inherent curvature of the DNA helix, likely play roles in the

increased affinity and specificity.^{8c,10b} Furthermore, examination of the isopotential surfaces of **H_z** and **H_p** indicates that favorable aromatic stacking interactions may be altered between the two different heterocyclic systems. Thus, by going from the 5-membered heterocyclic system of **H_p** to the fused 6-5 system of hydroxybenzimidazole **H_z**, changes associated with the indirect readout of the DNA-minor groove may be responsible for the evident changes in affinity and specificity.

Polyamides incorporated with multiple **H_z/P_y** pairs are moderately effective at differentiating between DNA sequences with multiple A,T sites, the specificity appearing to have a high degree of DNA sequence dependence. For example, polyamide **3** binds its match site 5'-AGTACT-3' with high fidelity. However, polyamide **4**, which is designed to target the site 5'-AGATCT-3' does so with virtually no specificity. Similarly, only modest specificity for the 5'-AGATCT-3' sequence is noted with multiple **H_p/P_y** pairs – an effect that may be attributed to the different DNA microstructure of the 5'-GAT-3' step in comparison to the 5'-GTA-3' step. This inherent sequence bias is evident upon

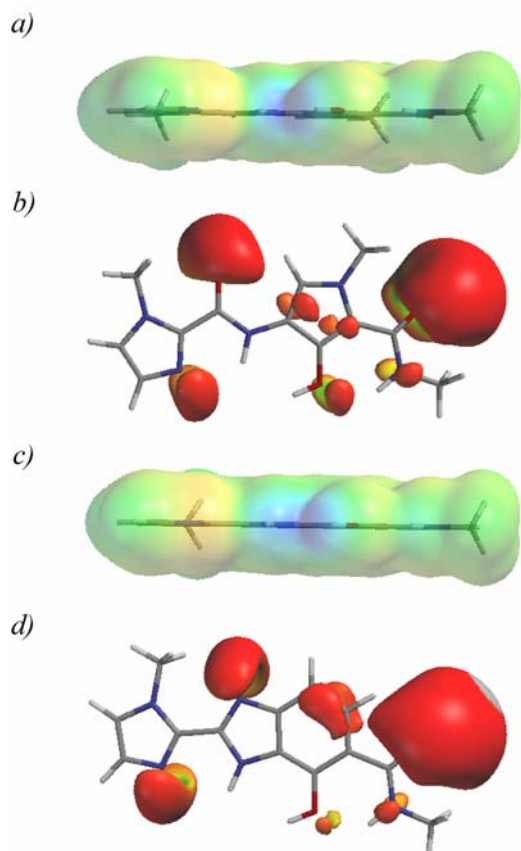


Figure 6.11. Calculated electronic surfaces of Im-Hp and Im-Hz dimeric units. *Ab initio* calculations done using 6-31G* basis set. (a) Electronic surface presented to the floor of the DNA minor groove by Im-Hp dimer. (b) Isopotential surface of the Im-Hp dimer. (c) Electronic surface presented to the floor of the DNA minor groove by Im-Hz dimer. (d) Isopotential surface of the Im-Hz dimer. Positive potentials are indicated by blue while negative potentials are indicated by red.

examination of the binding affinities of polyamides containing multiple symmetrical **Py/Py** pairs. The A,T degenerate **Py/Py** pair binds the 5'-AGTACT-3' sequence with a substantially higher affinity than the 5'-AGATCT-3' sequence, demonstrating the significant effect of DNA sequence on polyamide binding affinity.

Hydroxybenzimidazole/Benzimidazole Pair (Hz/Bi). The **Hz/Bi** pair presents the same functionality to the DNA-minor groove as the **Hz/Py** pair. While the functionality directed at the minor groove is the same, the **Hz/Bi** pair is the first example of a 6-5/6-5 bicyclic ring-ring pair. Accordingly, the overall geometry and electronics of the polyamide containing the **Hz/Bi** pair are substantially different (Figure 6.12). Like the **Hz/Py**

pair, the **Hz/Bi** pair is capable of discriminating between A,T Watson-Crick base pairs such that **Hz/Bi** codes for T·A and **Bi/Hz** codes for A·T. The affinity of the **Hz/Bi** pair is significantly higher than the **Hz/Py** pair, albeit demonstrates lower selectivity. Polyamides containing two fused benzimidazole analogues have substantially lower

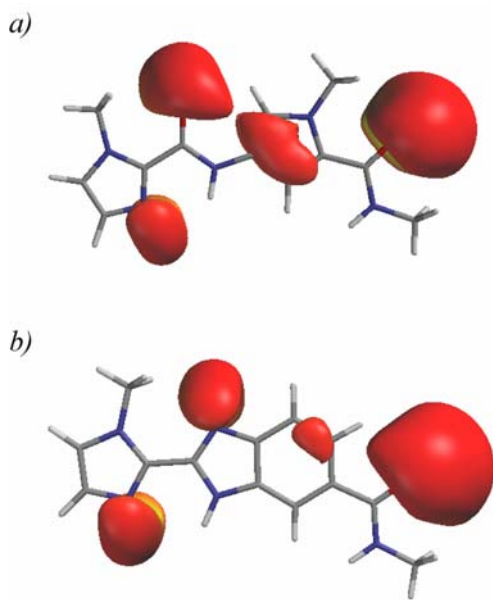


Figure 6.12. Calculated isopotential surfaces of Im-Py and Im-Bi dimeric units. *Ab initio* calculations done using 6-31G* basis set. (a) Isopotential surface of the Im-Py dimer. (b) Isopotential surface of the Im-Bi dimer. Positive potentials are indicated by blue while negative potentials are indicated by red.

conformational freedom and a higher degree of preorganization, favorably orienting the polyamides in the proper orientation to bind the DNA. Further, the benzimidazole moiety has a greater aromatic surface and hydrophobicity that may alter both the DNA-ligand van der Waals interactions, and the intrapolyamide π -stacking. The lower degree of curvature of the **H_z/B_i** pair may also allow for a more intimate polyamide/DNA interaction, contributing to the noted increase in binding affinity. These changes in structure, while providing a marked increase in affinity, do come at the

cost of lowered specificity. The lowered selectivity of the **H_z/B_i** pairing may be attributed to the greater rigidity of the system. More specifically, the lack of conformational freedom of the fused heterocyclic pairs may lower the sensitivity of the reading face of the polyamide, making it difficult to adjust to the small structural changes in the minor groove microenvironment that are the source of specificity.

Conclusion.

Previously, we have relied on hydroxypyrrole **H_p** as a T-specific recognition element, differentiating between T and A when paired with pyrrole **P_y**, such that **H_p/P_y** codes for T•A and **P_y/H_p** codes for A•T. In an effort to find a replacement for **H_p**, we

have examined the DNA recognition properties of polyamides containing **H_z/P_y** and **H_z/B_i** pairs. It was determined that the **H_z/P_y** pair is an effective and stable replacement for the previously reported **H_p/P_y** pair, successfully differentiating A,T base pairs at high affinity and specificity. It was also determined that multiple **H_z/P_y** pairs are moderately effective at discriminating multiple A,T sequences. Also, we report the use of the **H_z/B_i** pair, the first example of a 6-membered/6-membered ring-ring pair that is also capable of discriminating T·A from A·T base pairs. What is remarkable is that we are now on a pathway to constructing oligomers for minor groove recognition that are not based on pyrrole-carboxamides. We visualize a time when sequence specific designed oligomers may no longer be polyamides but are still based on the concept of pairing rules for DNA recognition. Future work will focus on the cellular uptake properties of these novel oligomers and their use in biological systems.

Acknowledgments.

We thank the National Institutes of Health for grant support, Caltech for a James Irvine Fellowship to R.M.D., the Parsons Foundation for a fellowship to M.A.M, and the NSF for a fellowship to S.F. We also thank C. Brindle for aid in preparing necessary intermediates.

Experimental.

General. N,N-dimethylformamide (DMF), N,N-diisopropylethylamine (DIEA), thiophenol (PhSH), N,N-dimethylaminopropylamine (Dp), triethylamine (TEA), and thiourea were purchased from Aldrich. Boc- β -alanine-(4-carboxylaminomethyl)-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. Precoated silica gel plates 60F₂₅₄ for TLC and silica gel 60 (40 μ m) for flash chromatography were from Merck. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [γ -³²P]triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) and DNase I (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA polymerase was from Perkin-Elmer and was used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d*₆ or CDCl₃ 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.

Monomer Synthesis. Compounds **7-13** and **19** were synthesized according to previously published protocols.³⁴⁻³⁶

Methy 7-methoxy-2-(1-methylimidazol-2-yl)benzimidazole-6-carboxylate (Im-HzOMe-OMe 14). A mixture of **13** (0.2 g, 1.02 mMol), Im-COCCl₃, (345 mg, 1.53 mMol), DIEA (132 mg, 178 μL, 1.02 mMol), and DMAP (25 mg, 204 μMol) in EtOAc (5 mL) was stirred for 12 h at 60 °C, over which time a precipitate formed. The reaction was cooled to room temperature, filtered, and washed with cold Et₂O. The off-white solid was collected and dissolved in AcOH (5 mL). The mixture was then heated to 90 °C for 6 h. The solvent was then removed by rotoevaporation and the remaining white solid dried under high vacuum to provide **14** (210 mg, 72% Yield). TLC (4:1 EtOAc/Hex) *R_f* 0.4; ¹H NMR (DMSO-*d*₆) 7.55 (s, 1H), 7.53 (s, 1H) 7.43 (s, 1H), 7.13 (s, 1H), 4.36 (s, 3H), 4.16 (s, 3H), 3.78 (s, 3H); ¹³C (DMSO-*d*₆) 166.4, 151.2, 143.8, 138.9, 136.9, 134.6, 128.3, 125.4, 125.3, 113.9, 105.1, 60.9, 51.7, 35.1; EI-MS *m/e* 286.107 (M⁺ calcd. for 286.107 C₁₄H₁₄N₄O₃).

7-methoxy-2-(1-methylimidazol-2-yl)benzimidazole-6-carboxylic acid (**Im-HzOMe-OH 16**). A mixture of **14** (200 mg, 699 μ Mol), MeOH (3 mL), and 1N NaOH (4 mL) was stirred at 35 °C for 3 h. The methanol was removed in vacuo and the mixture was taken to pH 2 using 1N HCl, upon which time a white precipitate formed. The mixture was poured into a 50 mL Falcon tube and spun down in a centrifuge (10 min x 14,000 rpm). The tube was decanted, leaving a white solid that was dried under high vacuum to provide **16** (165 mg, 87% Yield). TLC (3:2 EtOAc/Hex, 10% AcOH) R_f 0.4; ^1H NMR (DMSO- d_6) 7.58 (s, 1H), 7.55 (s, 1H) 7.46 (s, 1H), 7.16 (s, 1H), 4.33 (s, 3H), 4.16 (s, 3H); ^{13}C (DMSO- d_6) 167.4, 151.0, 143.6, 138.6, 136.8, 135.2, 127.8, 125.7, 125.3, 115.1, 105.2, 61.0, 35.2; EI-MS m/e 272.260 (M^+ calcd. for 272.260 $\text{C}_{13}\text{H}_{12}\text{N}_4\text{O}_3$).

Methyl 7-methoxy-2-(1-methyl-4-nitroimidazol-2-yl)benzimidazole-6-carboxylate (**NO₂-Im-HzOMe-OMe 15**). Diamine **13** (0.5 g, 2.54 mmol), NO₂-Im-OH **19** (480 mg, 2.80 mmol), HBTU (1 g, 2.66 mmol), DIEA (362 mg, 488 μ L, 2.80 mmol), and DMF (7 mL) were stirred for 2 days at room temperature. The mixture was then added to a 50 mL Falcon tube containing water (20 mL), resulting in a precipitate. The Falcon tube was centrifuged (10 min x 14,000 rpm) and the mother liquor decanted, leaving a tan solid that was dried under high vacuum. The solid was then dissolved in AcOH (8 mL) and heated to 90 °C with stirring. It is noteworthy that the solid was not completely soluble in AcOH. The reaction was stirred for 6 hours and the precipitate that was present was filtered over a fine fritted funnel. The solid was washed with Et₂O and dried under high vacuum to provide **15** (481 mg, 57% Yield) as a powdery yellow solid. TLC (3:2 EtOAc/Hex) R_f 0.5; ^1H NMR (DMSO- d_6) 8.68 (s, 1H), 7.60 (d, $J = 8.4$ Hz, 1H) 7.18 (d, $J = 8.4$ Hz, 1H), 4.38 (s, 3H), 4.24 (s, 3H), 3.78 (s, 3H); ^{13}C (DMSO- d_6) 166.2, 147.0,

142.1, 138.2, 137.7, 133.4, 132.7, 128.1, 117.5, 109.5, 107.3, 61.8, 52.1, 36.5; EI-MS m/e 331.092 (M^+ calcd. for 331.092 C₁₄H₁₃N₅O₅).

Methyl 2-{4-[(*tert*-butoxy)carbonylamino]-1-methylimidazol-2-yl}-7-methoxybenzimidazole-6-carboxylate (**Boc-Im-HzOMe-OMe 17**) A mixture of **15** (400 mg, 1.21 mmol), DIEA (400 mg, 536 μ L, 3.08 mmol), Pd/C (50 mg) and DMF (5 mL) was placed in a Parr apparatus and hydrogenated (500 psi) for 1.5 h at ambient temperature. The mixture was removed from the Parr apparatus and (Boc)₂O (396 mg, 1.82 mmol) was added. The mixture was then stirred for 8 h at 50 °C. The solvent was removed in vacuo, followed by column chromatography of the brown residue (3:2 EtOAc/Hex) to provide **17** as a thin film. The thin film was treated with hexanes and then the solvent was removed by rotoevaporation, followed by drying under high vacuum to provide **17** as a white solid (228 mg, 47% Yield). TLC (3:2 EtOAc/Hex) R_f 0.6; ¹H NMR (DMSO-*d*₆) 9.55 (s, 1H), 7.75 (d, J = 8.7 Hz, 1H) 7.60 (d, J = 8.7 Hz, 1H), 7.21 (s, 1H) 4.27 (s, 3H), 3.81 (s, 3H), 3.70 (s, 3H), 1.44 (s, 9H); ¹³C (DMSO-*d*₆) 165.8, 151.0, 147.0, 142.1, 137.9, 137.0, 133.3, 132.4, 127.8, 117.3, 109.4, 107.2, 85.9, 61.7, 52.0, 33.5, 28.2; EI-MS m/e 401.170 (M^+ calcd. for 401.170 C₁₄H₁₃N₅O₅).

2-{4-[(*tert*-butoxy)carbonylamino]-1-methylimidazol-2-yl}-7-methoxybenzimidazole-6-carboxylic acid (**Boc-Im-HzOMe-OH 18**). A mixture of **17** (200 mg, 498 μ Mol), 4N KOH (6 mL) and MeOH (4 mL) was stirred at 60 °C for 4 h. The MeOH was removed by rotoevaporation and the pH carefully 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 **18** (166 mg, 86% Yield) as a fine white solid. TLC (3:2 EtOAc/Hex, 10% AcOH) R_f 0.65; ¹H NMR (DMSO-*d*₆)

9.45 (s, 1H), 7.54 (d, $J = 8.7$ Hz, 1H) 7.23 (s, 1H), 7.14 (d, $J = 8.7$ Hz, 1H), 4.33 (s, 3H), 4.09 (s, 3H), 1.44 (s, 9H); ^{13}C (DMSO- d_6) 166.7, 150.8, 146.7, 142.0, 135.9, 137.0, 133.3, 132.4, 127.4, 117.3, 109.4, 107.3, 86.0, 61.8, 33.6, 28.2; EI-MS m/e 387.154 (M^+ calcd. for 387.154 $\text{C}_{18}\text{H}_{21}\text{N}_5\text{O}_5$).

Polyamide Synthesis. Polyamides 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.³⁶

Im-Im-**H_z**-Py- γ -Im-Py-Py-Py- β -Dp (**1**): (Boc-Im-**H_z**OMe-OH) (34 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 H_2N -Py- γ -Im-Py-Py-Py- β -Pam resin. Coupling was allowed to proceed for 12 h at room temperature. After Boc-deprotection, the terminal imidazole residue was incorporated using Im-COCCl₃. Im-COCCl₃ (67 mg, 295 μmol), DIEA (23 mg, 31 μl , 177 μmol) and DMF (400 μl) were added to the reaction vessel containing H_2N -Im-**H_z**OMe-Py- γ -Im-Py-Py-Py- β -Pam resin. Coupling was allowed to proceed for 2 h at 37 °C, and determined complete by analytical HPLC. The resin-bound polyamide was then washed with DCM and subjected to the cleavage, O-methyl deprotection and purification protocol described below to provide Im-Im-**H_z**-Py- γ -Im-Py-Py-Py- β -Dp (**1**) (1.1 mg, 3.1% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1233.56 ($\text{M}+\text{H}$ calcd. for 1233.56 $\text{C}_{58}\text{H}_{69}\text{N}_{22}\text{O}_{10}$).

Im-Im-Py-Py- γ -Im-**H_z**-Py-Py- β -Dp (**2**): (Boc-Im-**H_zOMe**-OH) was incorporated as described for **1**. The polyamide was cleaved from resin and treated as described in the deprotection protocol below to provide (**2**) (0.9 mg, 2.5 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1233.55 (M+H calcd. for 1233.56 C₅₈H₆₉N₂₂O₁₀).

Im-**H_z**-Py-Py- γ -Im-**H_z**-Py-Py- β -Dp (**3**): (Im-**H_zOMe**-OH) (25 mg, 88.5 μ mol) was incorporated by activation with HBTU (32 mg, 84 μ mol), DIEA (23 mg, 31 μ l, 177 μ mol) and DMF (250 μ l). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H₂N-Py-Py- γ -Im-**H_z**-Py-Py- β -Pam resin. 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-**H_z**-Py-Py- γ -Im-**H_z**-Py-Py- β -Dp (**3**) (0.7 mg, 1.9 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1242.56 (M+H calcd. for 1242.55 C₆₀H₆₈N₂₁O₁₀).

Im-Py-**H_z**-Py- γ -Im-Py-**H_z**-Py- β -Dp (**4**): (Boc-Py-**H_zOMe**-OH) (34 mg, 88.5 μ mol) was incorporated by activation with HBTU (32 mg, 84 μ mol), DIEA (23 mg, 31 μ l, 177 μ mol) and DMF (250 μ l). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H₂N-Py- β -Pam resin. Coupling was allowed to proceed for 12 h at room temperature. After Boc-deprotection, the additional **Im**, γ , and **Py** units were incorporated as previously described. The second Boc-Py-**H_zOMe**-OH unit was activated as described above and added to the reaction vessel containing H₂N-Py- γ -Im-Py-**H_z**-Py- β -Pam resin. Coupling was allowed to

proceed for 12 h at room temperature. After Boc-deprotection, the terminal imidazole residue was added as described for **1**. The resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide Im-Py-**H_z**-Py- γ -Im-Py-**H_z**-Py- β -Dp (**4**) (1.1 mg, 3.0 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1242.55 (M+H calcd. for 1242.55 C₆₀H₆₈N₂₁O₁₀).

Im-Im-**H_z**-Py- γ -Im-**Bi**-Py-Py- β -Dp (**5**): (Boc-Im-**Bi**-OH) (32 mg, 88.5 μ mol) was incorporated by activation with HBTU (32 mg, 84 μ mol), DIEA (23 mg, 31 μ l, 177 μ mol) and DMF (250 μ l). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H₂N-Py-Py- β -Pam resin. Coupling was allowed to proceed for 12 h at room temperature. After Boc-deprotection, the additional **Im**, γ , and **Py** units were incorporated as previously described. [Ref] The Boc-Im-**H_zOMe**-OH residue and the terminal imidazole residue were incorporated as described for **1**. The resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide Im-Im-**H_z**-Py- γ -Im-**Bi**-Py-Py- β -Dp (**5**) (1.1 mg, 3.0 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1226.55 (M+H calcd. for 1226.54 C₅₉H₆₆N₂₂O₉).

Im-Im-**Bi**-Py- γ -Im-**H_z**-Py-Py- β -Dp (**6**): (Boc-Im-**H_zOMe**-OH) and (Boc-Im-**Bi**-OH) were incorporated as described in **1** and **5**. The terminal imidazole residue was incorporated as described in **1**. Upon completion of the synthesis, the resin-bound polyamide was then washed with DCM and treated as described in the deprotection

protocol below to provide Im-Im-**Bi**-Py- γ -Im-**H_z**-Py-Py- β -Dp (**6**) (1.3 mg, 3.5 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1226.54 (M+H calcd. for 1226.54 C₅₉H₆₆N₂₂O₉).

Deprotection of the *O*-Methyl-Protected Polyamides. All of the above polyamides were cleaved from resin, purified, deprotected and subject to further purification using the following general procedure. Upon completion of solid phase synthesis, Dp (500 μ L) 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 (**-H_zOMe-**) as a dry solid. The polyamides were then dissolved in DMF (200 μ l) and added to a suspension of sodium hydride (40 mg, 60% oil dispersion) and thiophenol (200 μ l) in DMF (400 μ l) 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 **H_z**-containing polyamides.

Footprinting Experiments. Plasmids pDHN1 and pDEH10 were constructed and 5'-radiolabeled as previously described.^{26, 32} DNase I footprint titrations were performed according to standard protocols.¹

References:

- [1] Trauger, J. W.; Dervan, P. B., *Methods in Enzymology* **2001**, 340, 450-466.
- [2] Pandolfi, P. P., *Oncogene* **2001**, 20, (24), 3116-3127.
- [3] Darnell, J. E., *Nature Reviews Cancer* **2002**, 2, (10), 740-749.
- [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] Arcamone, F.; Nicoletti, V.; Penco, S.; Orezzi, P.; Pirelli, A., *Nature* **1964**, 203, (494), 1064.
- [6] 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.
- [7] Pelton, J. G.; Wemmer, D. E., *Proceedings of the National Academy of Sciences of the United States of America* **1989**, 86, (15), 5723-5727.
- [8] Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001**, 9, (9), 2215-2235.
- [9] Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology* **2003**, 13, (3), 284-299.
- [10] Dervan, P. B.; Burli, R. W., *Current Opinion in Chemical Biology* **1999**, 3, (6), 688-693.
- [11] Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature* **1997**, 387, (6629), 202-205.
- [12] 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, (22), 12890-12895.
- [13] Janssen, S.; Durussel, T.; Laemmli, U. K., *Molecular Cell* **2000**, 6, (5), 999-1011.
- [14] Ansari, A. Z.; Mapp, A. K., *Current Opinion in Chemical Biology* **2002**, 6, (6), 765-772.
- [15] Coull, J. J.; He, G. C.; Melander, C.; Rucker, V. C.; Dervan, P. B.; Margolis, D. M., *Journal of Virology* **2002**, 76, (23), 12349-12354.
- [16] Dudouet, B.; Burnett, R.; Dickinson, L. A.; Wood, M. R.; Melander, C.; Belitsky, J. M.; Edelson, B.; Wurtz, N.; Briehn, C.; Dervan, P. B.; Gottesfeld, J. M., *Chemistry & Biology* **2003**, 10, (9), 859-867.
- [17] Oyoshi, T.; Kawakami, W.; Narita, A.; Bando, T.; Sugiyama, H., *Journal of the American Chemical Society* **2003**, 125, (16), 4752-4754.

- [18] Dervan, P. B., *Faseb Journal* **1999**, 13, (7), A1423-A1423.
- [19] Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., *Science* **1994**, 266, (5185), 646-650.
- [20] Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nature Structural Biology* **1998**, 5, (2), 104-109.
- [21] Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., *Science* **1998**, 282, (5386), 111-115.
- [22] Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., *Journal of the American Chemical Society* **2001**, 123, (21), 5044-5049.
- [23] 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, (4), 1088-1095.
- [24] 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, (3), 557-567.
- [25] Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B., *Helvetica Chimica Acta* **2002**, 85, (12), 4485-4517.
- [26] Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001**, 9, (1), 7-17.
- [27] Doss, R. M.; Marques, M. A.; Foister, S.; Dervan, P. B., *Chemistry & Biodiversity* **2004**, 1, (6), 886-899.
- [28] Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2003**, 11, (20), 4333-4340.
- [29] Ellervik, U.; Wang, C. C. C.; Dervan, P. B., *Journal of the American Chemical Society* **2000**, 122, (39), 9354-9360.
- [30] White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1999**, 121, (1), 260-261.
- [31] 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.
- [32] Melander, C.; Herman, D. M.; Dervan, P. B., *Chemistry-a European Journal* **2000**, 6, (24), 4487-4497.
- [33] Minehan, T. G.; Gottwald, K.; Dervan, P. B., *Helvetica Chimica Acta* **2000**, 83, (9), 2197-2213.

- [34] Briehn, C. A.; Weyermann, P.; Dervan, P. B., *Chemistry-a European Journal* **2003**, 9, (9), 2110-2122.
- [35] Renneberg, D.; Dervan, P. B., *Journal of the American Chemical Society* **2003**, 125, (19), 5707-5716.
- [36] Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1996**, 118, (26), 6141-6146.
- [37] Bailey, D. M.; Johnson, R. E.; Salvador, U. J., *Journal of Medicinal Chemistry* **1973**, 16, (11), 1298-1300.
- [38] Harbuck, J. W.; Rapoport, H., *Journal of Organic Chemistry* **1972**, 37, (23), 3618-&.
- [39] Nishiwaki, E.; Tanaka, S.; Lee, H.; Shibuya, M., *Heterocycles* **1988**, 27, (8), 1945-1952.
- [40] Herman, D. E. California Institute of Technology, Pasadena, 2001.
- [41] *Sparten Essential*, Wavefunction Inc: 2000.