

## Chapter 10

### *Targeting G-Tetrads in the DNA Minor Groove*

*The text of this chapter was taken in part from a manuscript coauthored with David M. Chenoweth, Raymond M. Doss, Eric J. Fechter and Professor Peter B. Dervan (Caltech) (Marques, M. A.; Chenoweth, D. M.; Doss, R. M.; Fechter, E. J. and Dervan, P. B. "Targeting G-Tetrads in the DNA Minor Groove" Journal of the American Chemical Society 2005 – In Preparation.*

**Abstract.**

Previously, cofacial asymmetric pairs consisting of the 5-membered heterocycles pyrrole (**Py**), hydroxypyrrole (**Hp**) and imidazole (**Im**) have been highly successful at coding for the four Watson-Crick base pairs in the DNA minor groove. More recently, the recognition of DNA by this class of molecules has evolved from polyamides consisting of 5-membered heterocyclic carboxamides to a new class of oligomers consisting predominantly of the 6-5 fused heterocyclic derivatives: benzimidazole (**Bi**), hydroxybenzimidazole (**Hb**), and imidazopyridine (**Ip**). These benzimidazole analogues are capable of mimicking their 5-membered counterparts and have demonstrated improved DNA recognition properties in a variety of sequence contexts. A series of oligomers containing the **Bi** and **Ip** ring systems were developed to target the G-tetrad sequence (5'-GGGG-3') in the DNA minor groove. The DNA of the G-tetrad is characterized by a wide, shallow, and undercurved minor groove. Previous attempts at recognizing the 5'-GGGG-3' sequence at high affinity were unsuccessful. However, the new series of **Bi** and **Ip** containing oligomers is capable of targeting the G-tetrad at high affinity and moderate specificity. The success of these new oligomers is attributed to a combination of hydrophobic and geometric properties.

## Introduction.

Aberrant gene expression is the cause for a large number of disease states. The ability to selectively manipulate genes, effectively reprogramming the cellular machinery at the most fundamental level, has substantial implications for modern medicine. One anti-gene approach involves the use of small synthetic molecules that have the ability, through rational design, to selectively target a DNA sequence of choice with high specificity. Of a large panel of DNA binding drugs, netropsin and distamycin have served as the starting scaffold for potent minor groove DNA binding ligands.<sup>2</sup>

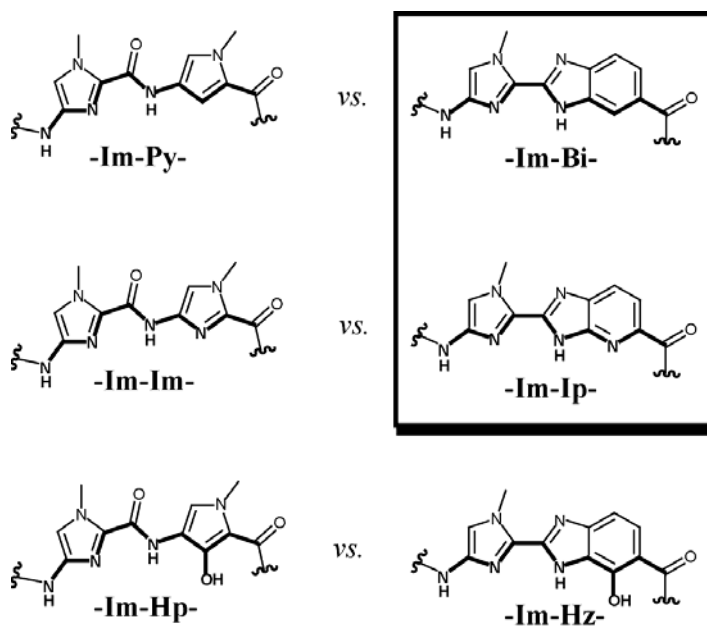
After many rounds of molecular evolution, polyamides were developed that recognize predetermined DNA sequences with high affinity and fidelity, mimicking the recognition profiles of naturally occurring DNA binding proteins.<sup>3-5</sup> Minor groove DNA recognition results from a variety of complex interactions, but the clearest modulation in recognition comes from the antiparallel co-facial pairings of heterocycles. The recognition face of the heterocycles has the potential to form favorable and unfavorable interactions with the Watson-Crick bases of the helix. For example, Net and Dis are known to bind poly A,T sequences due to their inability to accommodate the exocyclic amine of guanine, a negative steric interaction.<sup>6-9</sup> By analogy, a hydrogen bond acceptor should be able to selectively recognize the exocyclic amine of guanine and form a favorable hydrogen bonding interaction. This hypothesis was first successfully tested by incorporating a pyridyl group into the Dis scaffold, changing distamycin's inherent A,T selectivity to that of a G,C specific ligand.<sup>10-12</sup> Using these fundamental ideas originating from the unlinked polypyrrole scaffolds of *Net* and *Dis*, a polyamide:DNA

recognition code was developed, with specific heterocyclic ring pairs coding for a respective Watson Crick base pair. The pyrrole/pyrrole (**Py/Py**) pair is degenerate for A,T while imidazole paired against pyrrole as an (**Im/Py**) or (**Py/Im**) pair, codes for G•C and C•G respectively.<sup>13-15</sup> Finally, hydroxypyrrole paired against pyrrole as an (**Hp/Py**) or (**Py/Hp**) pair codes for T•A and A•T respectively.<sup>16-20</sup>

Since the development of these pairing rules based on the 5-membered heterocyclic carboxamides **Py**, **Im** and **Hp**, hundreds of DNA sequences have been successfully coded for.<sup>3, 5, 21</sup> Yet, due to the high degree of variability in DNA structure, there remain a variety of DNA sequences that prove difficult to target successfully. Furthermore, recent cell trafficking studies suggest that small alterations in ligand structure can have substantial effects on nuclear permeability.<sup>22-24</sup> To address targeting these problematic sequences and further explore the link between ligand structure and nuclear permeability, there has been an ongoing effort to develop novel heterocyclic elements with improved DNA recognition and cellular trafficking profiles. The scope of the search for new recognition elements originated with a small privileged library of 5-membered heterocyclic carboxamides and eventually moved to include 6-5 fused benzimidazole derivatives.<sup>25-28</sup> The benzimidazole derivatives are able to mimic their 5-membered counterparts with respect to the substituents presented to the DNA minor groove (Figure 10.1). Stemming from the work with the benzimidazole scaffold, a new series of DNA recognition guidelines were established. The benzimidazole/pyrrole (**Bi/Py**) pair mimics the **Py/Py** pair and is degenerate for A,T bases.<sup>29</sup> The hydroxybenzimidazole/pyrrole (**Hb/Py**) pair mimics the hydroxypyrrole/pyrrole (**Hp/Py**)

pair and selectively codes for T•A bases.<sup>30, 31</sup> Finally, the imidazopyridine/pyrrole (**Ip/Py**) pair mimics the (**Im/Py**) pair and codes for the G•C base pairing.<sup>30</sup>

The success of this new class of oligomers, comprising predominately fused



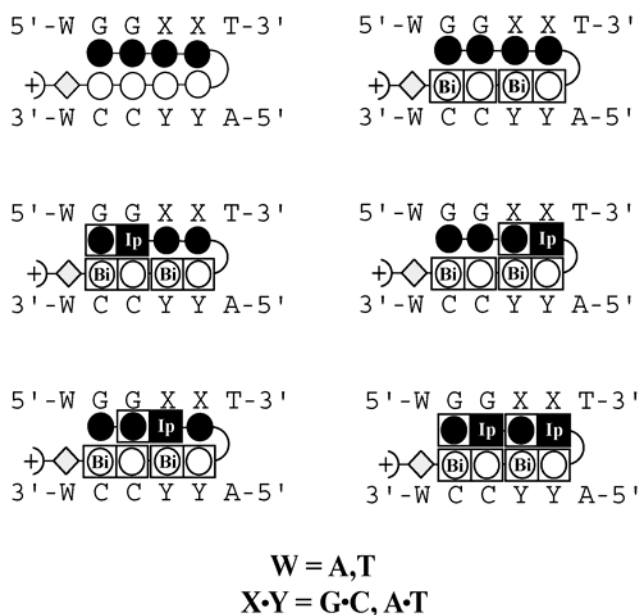
**Figure 10.1.** Traditional 5-membered heterocyclic carboxamides, imidazole (**Im**), pyrrole (**Py**) and hydroxypyrrole (**Hp**) in comparison to the benzimidazole analogue mimics, benzimidazole (**Bi**), imidazopyridine (**Ip**) and hydroxybenzimidazole (**HZ**). The recognition edge presented to the DNA minor groove is shown in bold.

benzimidazole derivatives, at binding selected DNA sequences with impressive affinity and fidelity suggested that they may be useful in targeting DNA sequences that have been shown to be problematic. One such sequence is a G-tetrad (5'-WGGGGW-3', W = A,T). Previous attempts at recognizing poly-guanine tracts using traditional **Py** and

**Im** containing polyamides met with little success.<sup>32, 33</sup> Quantitative DNase I footprinting titration showed polyamides designed to target the 5'-WGGGGW-3' sequence bound with very low affinity ( $K_a \sim 10^7 \text{ M}^{-1}$ ). One explanation for the substantial loss in affinity was an inherent lack of complementarity caused by a mismatch in curvature between the DNA helix and the polyamide ligand.<sup>32, 34, 35</sup> Such a lack of complementarity could result in loss of favorable hydrogen bonding between the contiguous exocyclic amines of guanine and the N(3) nitrogen of the imidazole rings. The small collection of structural

data on poly-guanine DNA sequences suggests that the helix is highly A-form with a substantially less curved, more shallow and wider minor groove.<sup>36-39</sup>

In an attempt to improve recognition of the G-tetrad (5'-GGGG-3') DNA, a series



**Figure 10.2.** Ball and stick models of oligomers in the context of a variable DNA sequence. Examination of flanking sequence effects on oligomer binding are tested by varying W to both A and T Watson-Crick bases. Examination of mismatch specificity is tested by varying X•Y in the core binding sequence to both G•C and A•T respectively. Imidazole is indicated by a dark circle. Pyrrole is indicated by a white circle. Imidazopyridine-imidazole (**Ip-Im**) and benzimidazole-pyrrole (**Bi-Py**) dimers are indicated by dark and light rectangles, respectively.  $\beta$ -alanine is depicted as a diamond and  $\gamma$ -aminobutyric acid is depicted as a half circle.

of novel oligomers (**1-6**)

designed to target the poly-guanine DNA sequence, and that

vary in heterocycle content, were

prepared. Oligomers vary by

incorporation of benzimidazole-

pyrrole (**Bi-Py**) dimers and

placement of imidazopyridine-

imidazole (**Ip-Im**) dimers

(Figure 10.2). While each

oligomer does present the same

functionality to the DNA minor

groove as shown in Figure 10.1,

each oligomer displays different

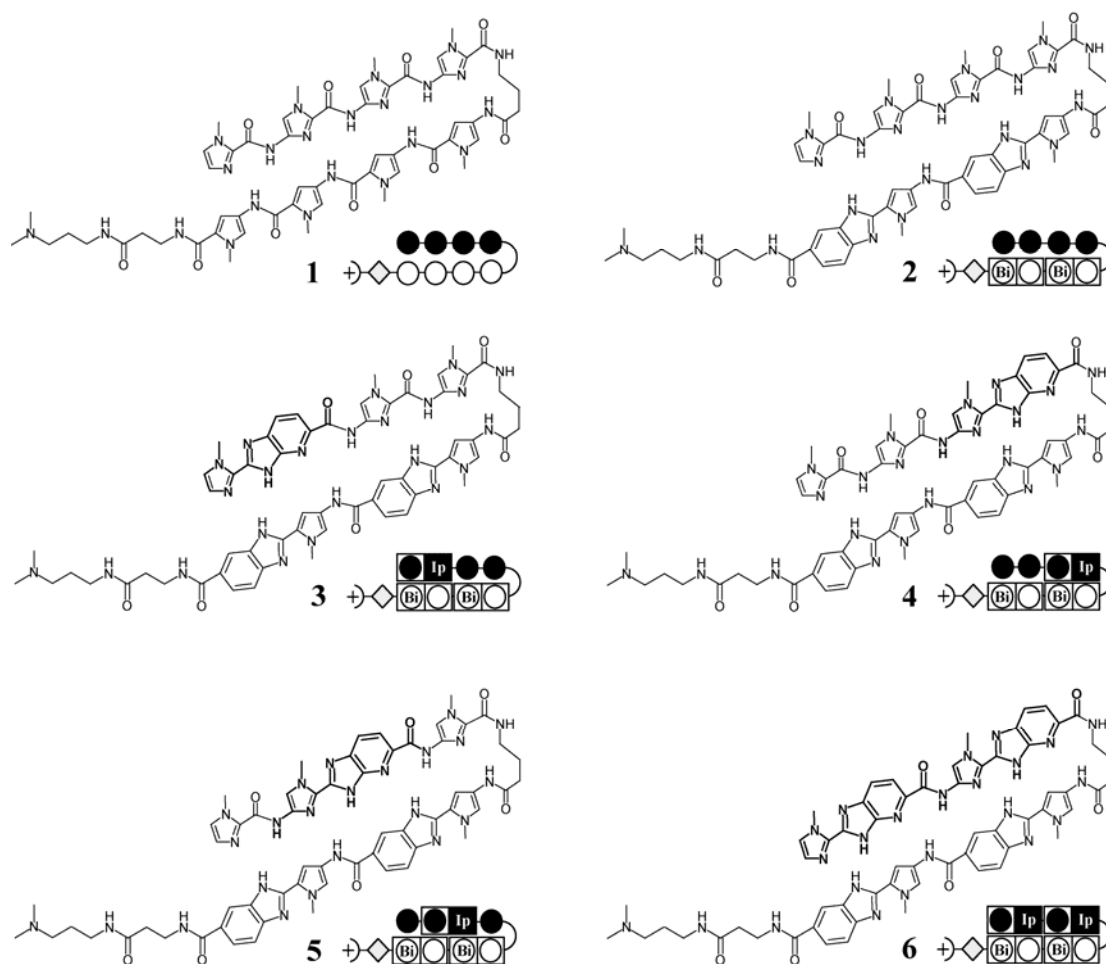
geometric and electronic profiles

due to heterocycle composition

(Figure 10.3). The ability of

each oligomer to target the 5'-GGGG-3' sequence specifically in the context of other

mismatch sequences was examined using quantitative DNase I footprinting titrations.



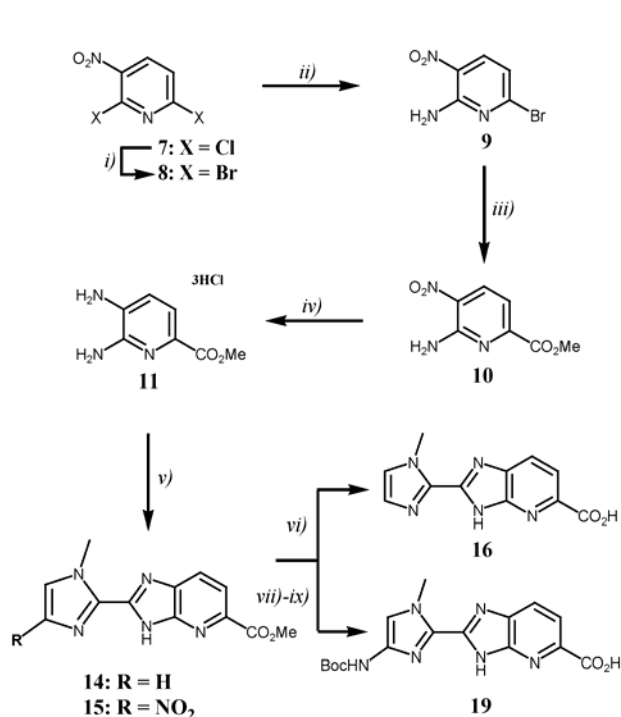
**Figure 10.3.** Chemical structure of oligomers **1-6** shown with imidazopyridine-imidazole (**Ip-Im**) dimers in bold. With each chemical structure is the accompanying ball and stick model for the oligomer. Imidazopyridine-imidazole (**Ip-Im**) and benzimidazole-pyrrole (**Bi-Py**) dimers are indicated by dark and light rectangles, respectively.  $\beta$ -alanine is depicted as a diamond and  $\gamma$ -aminobutyric acid is depicted as a half circle.

Molecular modeling was used to provide insight into the oligomer architecture and to collaborate the thermodynamic results provided by footprinting titration reported here.

## Results.

**Heterocycle Synthesis.** Dimeric units Im-*Ip*-OH (**16**) and Boc-Im-*Ip*-OH (**19**) were synthesized starting from commercially available 2,6-dichloro-3-nitropyridine

(7) (Figure 10.4). Nitropyridine (7) was converted to its dibromo derivative (8) by halogen exchange at both *ortho*-positions using a combination of HBr in acetic acid and



**Figure 10.4.** Synthesis of imidazopyridine-imidazole dimers Im-*Ip*-OH (**16**) and Boc-Im-*Ip*-OH (**19**). i) 30% HBr in AcOH, NaBr, H<sub>2</sub>SO<sub>4</sub>; ii) NH<sub>3</sub> in MeOH; iii) Pd(OAc)<sub>2</sub>, MeOH, CO; iv) H<sub>2</sub>, Pd/C, DMF; v) Im-CHO (**12**) or NO<sub>2</sub>-Im-CHO (**13**), NO<sub>2</sub>Ph; vi) 4N KOH, MeOH; vii) H<sub>2</sub>, Pd/C, DMF; viii) (Boc)<sub>2</sub>O, DIEA, DMAP, DMF; ix) 4N KOH, MeOH.

and complete conversion of starting material can be accomplished by equipping the reaction flask with a hydrogen balloon.

From **11**, a variety of conditions can be employed to effect the transformation to fused bicyclic derivatives such as **14**. First, heterocyclic carboxylic acids can be activated and mixed with their corresponding diamine coupling partners to form the initial amide bond, followed by cyclodehydration to provide the fused benzimidazole

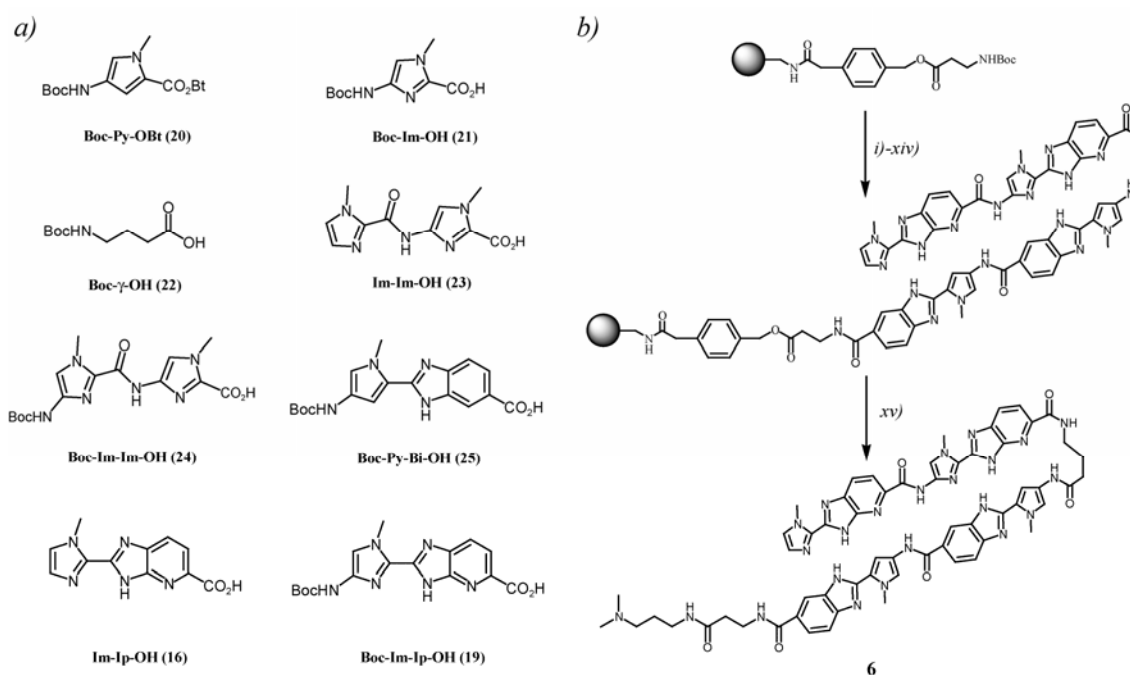
anhydrous HBr.<sup>40</sup> Regioselective amination of (**8**) using a saturated solution of NH<sub>3</sub> in MeOH provided (**9**).<sup>30</sup> Continuous bubbling of anhydrous NH<sub>3</sub> into the reaction vessel provides improved yields. Palladium-catalyzed carbonylation of (**9**) provided (**10**) in moderate yield as previously described.<sup>41, 42</sup> Reduction of **10** using Pd/C in the presence of hydrogen provided the corresponding diamine (**11**) cleanly and in near quantitative yields. It is noteworthy that increased hydrogen pressure (~500 psi) is not necessary



system.<sup>29, 31</sup> This initial approach did not allow access to the desired imidazopyridine intermediate **14**. A second approach, mixing diaminopyridine intermediate (**11**) and commercially available imidazole aldehyde (**12**) in the presence of an oxidant such as FeCl<sub>3</sub> or benzoquinone was only moderately successful, providing **14** in low yield following chromatography.<sup>43</sup> Ultimately, mixing **12** and **11** in refluxing nitrobenzene overnight provided **14** in excellent yield without the need for chromatography.<sup>44-46</sup> **14** was then saponified using a mixture of KOH (4M) in MeOH under heating to provide the final Im-Ip-OH dimer (**16**).

Nitroimidazole (**13**) was synthesized from **12** using a mixture of oleum and neat red fuming nitric acid.<sup>47</sup> All attempts at nitration using a variety of other conditions, such as 70% nitric acid or 90% nitric acid with sulfuric acid in acetic anhydride over a broad range of temperatures (-30-60 °C) were unsuccessful. Nitronium reagents were also unsuccessful. **13** was then added to a mixture of **11** in nitrobenzene and refluxed at 140 °C open to the atmosphere overnight to provide NO<sub>2</sub>-Im-Ip-OMe (**15**) cleanly. **15** was reduced using Pd/C in the presence of hydrogen to provide **17**. The amine (**17**) was then Boc-protected using a mixture of Boc anhydride and DMAP in DMF to give **18**. It is noteworthy that the amine **17** is sufficiently unreactive and must be heated with excess (Boc)<sub>2</sub>O in the presence of DMAP for the reaction to proceed. Following a single chromatography, saponification of **18** provided the final dimer Boc-Im-Ip-OH (**19**).

**Oligomer Synthesis.** Oligomers **1-6** were synthesized using manual solid phase synthesis methodology on commercially available  $\beta$ -Ala-PAM resin as previously described.<sup>48</sup> Starting from base resin, monomeric and dimeric heterocyclic units were appended onto the resin in stepwise fashion using HBTU activation (Figure 10.5).



**Figure 10.5.** Solid phase synthesis of oligomers. a) heterocyclic and aliphatic monomers and dimers used to synthesize oligomers **1-6**. b) solid phase synthesis of oligomer **6**: i) 80% TFA in DCM; ii) Boc-Py-Bi-OH, HBTU, DIEA, DMF; iii) Ac<sub>2</sub>O, DIEA, DMF; iv) 80% TFA in DCM; v) Boc-Py-Bi-OH, HBTU, DIEA, DMF; vi) Ac<sub>2</sub>O, DIEA, DMF; vii) 80% TFA in DCM; viii) Boc- $\gamma$ -OH, HBTU, DIEA, DMF; ix) Ac<sub>2</sub>O, DIEA, DMF; x) 80% TFA in DCM; xi) Boc-Im-Ip-OH, HBTU, DIEA, DMF; xii) Ac<sub>2</sub>O, DIEA, DMF; xiii) 80% TFA in DCM; xiv) Im-Ip-OH, HBTU, DIEA, Couplings were allowed to proceed for several hours between 25-40 °C. Unreacted amines were acylated between coupling rounds using acetic anhydride. Deprotection of the Boc-protected amines was accomplished using 80% TFA in DCM. After completion of solid-phase synthesis, the resin was treated with dimethylaminopropylamine (Dp) and the oligomers were purified by reverse phase preparatory HPLC: Im-Im-Im-Im- $\gamma$ -Py-Py-

Py-Py- $\beta$ -Dp (1), Im-Im-Im-Im- $\gamma$ -PyBi-PyBi- $\beta$ -Dp (2), ImIp-Im-Im- $\gamma$ -PyBi-Py-Bi- $\beta$ -Dp (3), Im-Im-ImIp- $\gamma$ -PyBi-PyBi- $\beta$ -Dp (4), Im-ImIp-Im- $\gamma$ -PyBi-PyBi- $\beta$ -Dp (5), ImIp-ImIp- $\gamma$ -PyBi-PyBi- $\beta$ -Dp (6). Oligomers were characterized using MALDI-TOF spectrometry.

**DNA Affinity and Sequence Specificity.** Quantitative DNase-I-footprinting titrations were carried out for oligomers 1-6 on the PCR product of plasmid pEF16 (Figure 10.6). Plasmid pEF16 was constructed containing two designed match sites (5'-XGGGGT-3' X = A, T) and two mismatch sites (5'-AGGGAT-3' and 5'-AGAGGT-3') (Figure 10.4). The first two match sites 5'-TGGGGT-3' and 5'-AGGGGT-3' were

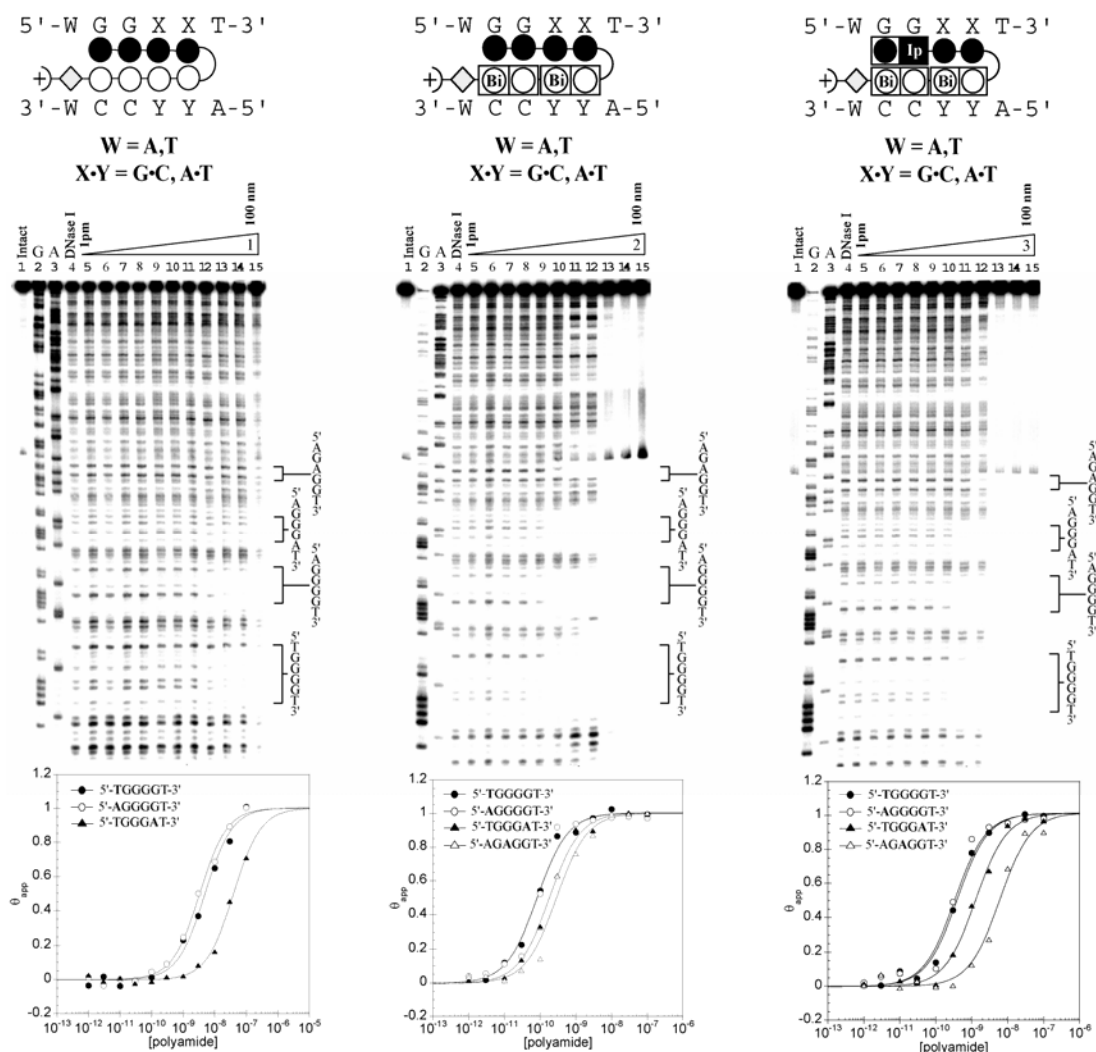


**Figure 10.6.** Sequence of plasmid pEF16 used for DNaseI footprinting titrations. The four designed binding sites are shown boxed with the variable base pair depicted in yellow.

constructed to determine if there was an energetic penalty associated with a 5'-AG-3' step. Previously, DNA sequences containing multiple 5'-AG-3' transitions have proven more difficult to target at high affinity and it has been shown that changes in flanking sequence can effect binding at a proximal site.<sup>49-51</sup> The second two designed binding sites, 5'-AGGGAT-3' and 5'-AGAGGT-3' were designed to elucidate the energetic penalty for the loss of a favorable hydrogen bond between the exocyclic amine of guanine and the lone pair nitrogen on the oligomer in question.

Control compound **1** bound both 4-G match sequences (5'-XGGGGT-3', X = A,T) with comparably low affinity ( $K_a \sim 10^8 \text{ M}^{-1}$ ), showing no bias for either site. **1** distinguished against mismatch sequences (5'-AGGGAT-3' and 5'-AGAGGT-3') with

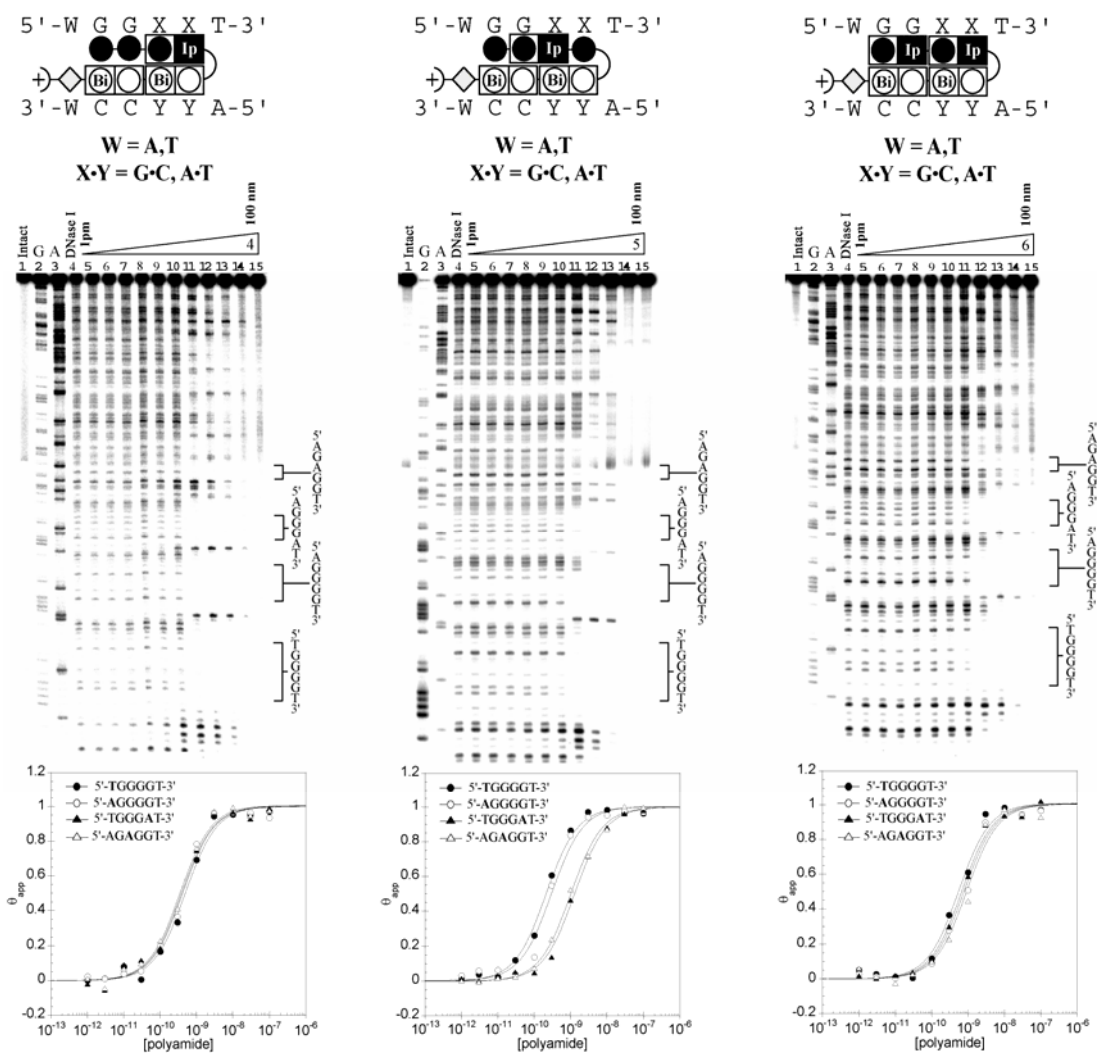
roughly 10-fold specificity (Figure 10.7). Oligomer **2** demonstrated a large increase in affinity ( $K_a \sim 10^{10} \text{ M}^{-1}$ ) as compared to **1** for both match sites but showed lower



**Figure 10.7.** DNase I footprinting titrations for control compound **1** and oligomers **2** and **3**. (Top) Ball and stick structure of each oligomer. Imidazopyridine-imidazole (**Ip-Im**) and benzimidazole-pyrrole (**Bi-Py**) dimers are indicated by dark and light rectangles, respectively.  $\beta$ -alanine is depicted as a diamond and  $\gamma$ -aminobutyric acid is depicted as a half circle. (Middle) DNase I footprinting titration shown with lanes from left to right: Intact, G and A sequencing lanes, DNase I control, lanes 5-15 oligomer concentrations 1pM to 100 nM. (Bottom) Quantitative isotherms depicting the affinity and specificity of each oligomer. Isotherms were generated using previously published methods.<sup>1</sup>

specificity (4-fold) over the designed mismatch sites. Oligomers **3** and **4** showed a

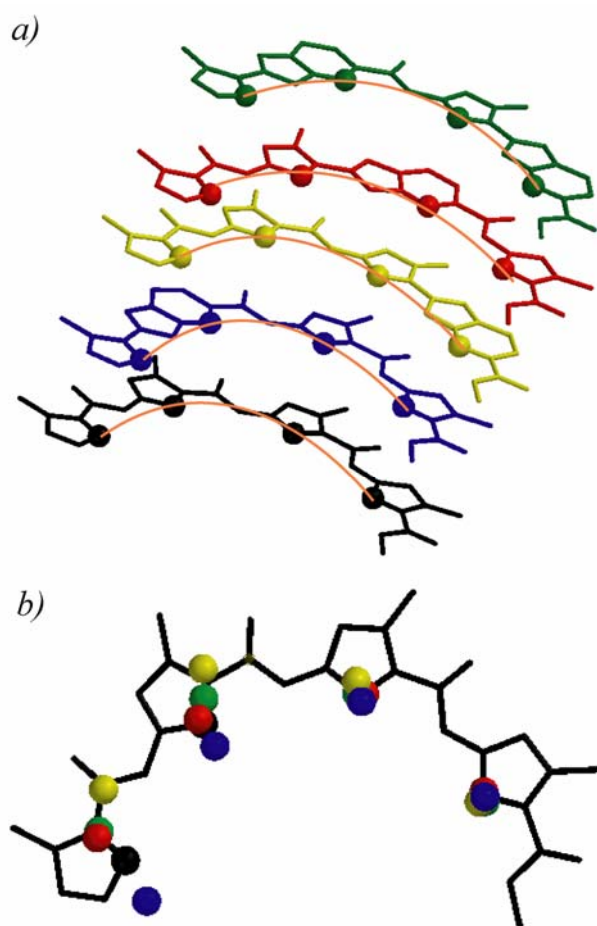
moderate increase in affinity ( $K_a \sim 10^9 \text{ M}^{-1}$ ) but demonstrated only minor selectivity over



**Figure 10.8.** DNase I footprinting titrations for oligomers **4-6**. (Top) Ball and stick structure of each oligomer. Imidazopyridine-imidazole (**Ip-Im**) and benzimidazole-pyrrole (**Bi-Py**) dimers are indicated by dark and light rectangles, respectively.  $\beta$ -alanine is depicted as a diamond and  $\gamma$ -aminobutyric acid is depicted as a half circle. (Middle) DNase I footprinting titration shown with lanes from left to right: Intact, G and A sequencing lanes, DNase I control, lanes 5-15 oligomer concentrations 1 pM to 100 nM. (Bottom) Quantitative isotherms depicting the affinity and specificity of each oligomer. Isotherms were generated using previously published methods.<sup>1</sup>

mismatch sites. Oligomer **5** demonstrated high affinity for the 4-G match sequences ( $K_a \sim \text{mid } 10^9 \text{ M}^{-1}$ ) with reasonable 5-fold selectivity over mismatch sequences (5'-AGGGAT-3' and 5'-AGGAGT-3') (Figure 10.8). **6** bound all designed sequences with similar affinity. Thermodynamic data for oligomers **1-6** is summarized in Table 10.1.

**Molecular Modeling.** Modeling calculations were performed with the *Spartan*



**Figure 10.9.** *Ab initio* modeling of oligomeric top strands. a) differential curvature of: Im-Im-Im-Im (black), Im-Im-ImIp (blue), ImIp-Im-Im (yellow), Im-ImIp-Im (red), and ImIp-ImIp (green). b) Overlay of oligomeric top strands with position of hydrogen bond accepting nitrogen shown. Color scheme is the same as described above.

*Essential* software package. *Ab initio* calculations were done using a Hartree-Fock model and a 6-31G\* polarization basis set. Four-ring subunits containing the sequences Im-Im-Im-Im, **ImIp**-Im-Im, Im-Im-**ImIp**, Im-**ImIp**-Im and **ImIp-ImIp** were constructed to examine respective overall ligand geometry and relative location of the nitrogen lone pair presented to the DNA minor groove (Figure 10.9). To visualize the broad structural differences between A-form and B-form DNA, two helices of sequence 5'-GAT**GGGG**TAC-3' were generated using the following parameters: A-form rise/base 2.55

angstroms and twist/base  $32.7^\circ$  and B-form rise/base 3.38 angstroms and twist/base  $36.0^\circ$  (Figure 10.10a). The same parameters were used to generate two A-form helices of sequence 5'-GGGG-3' and 5'GAGG-3' to demonstrate the minimal structural differences of the DNA minor groove by substitution of guanine for adenine (Figure 10.10b).

## Discussion.

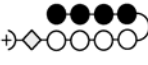
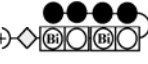
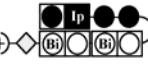
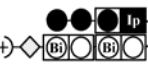


While classic Py/Im polyamides are capable of targeting a wide array of DNA sequences, there remain some sequences that, due to deviation from standard B-form DNA structure, are difficult to recognize with high affinity and specificity.<sup>4, 52-54</sup> Changes in the DNA microenvironment may result in helices that have substantially narrow minor groove widths, making 2:1 ligand binding thermodynamically unfavorable.<sup>50, 55</sup> In other cases, such as the polypurine 5'-GGGG-3' sequence, the DNA is shown to take on an A-form structure.<sup>36-39, 56</sup> In contrast to canonical B-form DNA, the A-form helix of the G-tetrad is characterized by a wide, shallow, and less curved minor groove (Figure 10.10a). With respect to helical parameters, the poly-G sequence shows a lower rise-per-residue and twist, while displaying a positive roll.<sup>36, 37</sup> The more accessible minor groove of the A-form DNA also results in a larger spine of hydration, possibly making binding the sequence less unfavorable.<sup>36, 57</sup>

Previously, Swalley et al. attempted to target both 5'-GGG-3' and 5'-GGGG-3' sequences using polyamides containing contiguous Py/Im pairs, such as compound **1**. One question posed in the later study was, does having high guanine content result in lowered affinities? A large number of exocyclic amines in the minor groove was theorized to present a significant steric block to an incoming polyamide and be thermodynamically unfavorable.<sup>58, 59</sup> However, Swalley et al. determined that sequence context was more important than sequence content. More specifically, a DNA sequence of 5'-GGCC-3', containing four guanine bases could be bound at high affinity, while the sequence 5'-GGGG-3', containing four contiguous guanine bases, could not.<sup>32</sup> The explanation offered for the substantial loss in affinity of the polyamide:DNA complex in

the case of the contiguous 4-G sequence was based on complex complementarity.

Swalley et al. proposed that the contiguous imidazole heterocycles of the polyamide Im-Im-Im- $\gamma$ -Py-Py-Py-Py- $\beta$ -Dp (**1**) were over curved with respect to the DNA helix, and not able to correctly align with the four contiguous exocyclic amines of the guanine bases.

Table 10.1. *G*-Tetrad Specific Oligomers  $K_a$  [ $M^{-1}$ ]<sup>a,b</sup>

Oligomer	5'-ATGGGGT-3'	5'-AAGGGGT-3'	5'-ATGGGAT-3'	5'-ATGAGGT-3'
<b>1</b> 	1.4 ( $\pm 1.0$ ) $\times 10^8$	2.6 ( $\pm 1.1$ ) $\times 10^8$	2.3 ( $\pm 0.8$ ) $\times 10^7$	< 1.0 $\times 10^7$
<b>2</b> 	1.9 ( $\pm 1.4$ ) $\times 10^{10}$	2.0 ( $\pm 1.1$ ) $\times 10^{10}$	4.8 ( $\pm 1.1$ ) $\times 10^9$	3.6 ( $\pm 0.9$ ) $\times 10^9$
<b>3</b> 	2.6 ( $\pm 0.5$ ) $\times 10^9$	2.9 ( $\pm 0.2$ ) $\times 10^9$	8.6 ( $\pm 2.1$ ) $\times 10^8$	2.5 ( $\pm 0.8$ ) $\times 10^8$
<b>4</b> 	2.6 ( $\pm 0.6$ ) $\times 10^9$	2.7 ( $\pm 0.4$ ) $\times 10^9$	2.6 ( $\pm 0.2$ ) $\times 10^9$	2.9 ( $\pm 0.1$ ) $\times 10^9$
<b>5</b> 	4.4 ( $\pm 0.9$ ) $\times 10^9$	4.1 ( $\pm 1.2$ ) $\times 10^9$	8.1 ( $\pm 2.3$ ) $\times 10^8$	8.2 ( $\pm 1.3$ ) $\times 10^8$
<b>6</b> 	1.6 ( $\pm 0.2$ ) $\times 10^9$	1.1 ( $\pm 0.4$ ) $\times 10^9$	1.3 ( $\pm 0.1$ ) $\times 10^9$	1.1 ( $\pm 0.1$ ) $\times 10^9$

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0.

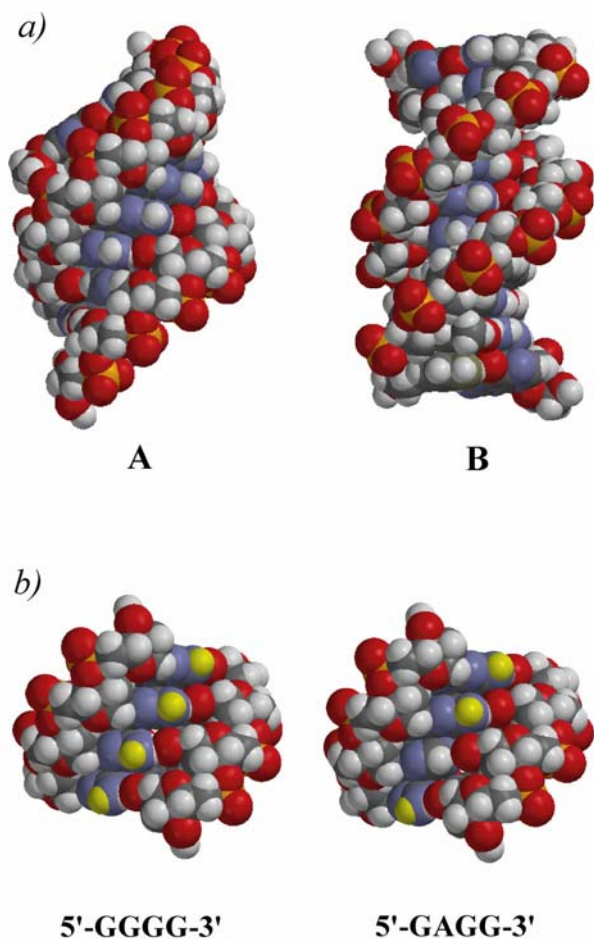
The combination of structural data on the 5'-GGGG-3' DNA tract and the previous work toward binding the sequence suggested that the G-tetrad could be specifically targeted at higher affinities by altering the ligand geometry. Recently, Renneberg et al. reported the utility of imidazopyridine/pyrrole pair (**Ip/Py**) as a functional imidazole/pyrrole (**Im/Py**) pair mimic.<sup>30</sup> A fused 6-5 bicycle, **Ip** is able to present the same recognition face to the DNA minor groove, coding for the exocyclic amine of guanine via hydrogen bond formation, while offering a significantly different geometry. Incorporating **Ip** into different positions within the oligomeric top strands, oligomers **3-6** were generated, each one possessing a unique geometry. Variation of the oligomer curvature, and subsequently the nitrogen lone-pair position, was envisioned to generate oligomers capable of hydrogen bonding to the exocyclic amines of the 5'-



GGGG-3' sequence more effectively (Figure 10.9). Furthermore, previous work by Briehn et al. and others demonstrated the utility of the benzimidazole (**Bi**) bicycle at being an effective high affinity replacement for pyrrole (**Py**).<sup>29</sup> By analogy, oligomers containing multiple imidazole/benzimidazole (**Im/Bi**) or imidazopyridine/benzimidazole (**Ip/Bi**) pairs were reasoned to display a relative gain in absolute affinity.

As reported by Swalley et al., compound **1**, containing four contiguous **Im/Py** pairs bound the G-tetrad sites with low affinity. Examination of the curvature and position of the N(3) groups presented to the DNA minor groove by the four contiguous imidazole (**Im**) carboxamides shows that the Im-Im-Im-Im- strand is significantly more curved than the remaining oligomers (**3-6**) included in this study (Figure 10.9). While the affinity for the 5'-GGGG-3' sites is low, compound **1** is able to discriminate against the 5'-GGGA-3' and 5'-GAGG-3' mismatch sites effectively. Of particular note, oligomer **2**, which maintains the four contiguous imidazole carboxamides in its top strand, while pairing them against two contiguous benzimidazole-pyrrole (**Bi-Py**) dimers, shows a significant increase in affinity of roughly 100-fold ( $K_a \sim 10^{10} \text{ M}^{-1}$ ) for the 5'-GGGG-3' match sites. However, oligomer **2** exhibits a minor 3-fold selectivity over the designed mismatches. This result seems to indicate that having an optimal orientation between the guanine exocyclic amines and the hydrogen bond acceptors on the oligomer is not the only factor governing recognition of the G-tetrad, and that the overall geometry and content of both strands of the oligomer play a substantial role in ligand affinity and specificity. Furthermore, this result implies that the poor affinity of compound **1**, is not completely due to the geometry of the four contiguous imidazole carboxamides (Im-Im-Im-Im-), a question that has been unanswered until now.

Oligomers **3-6**, incorporated with the imidazopyridine (**Ip**) ring system, and that maintain a bottom strand consisting of two **Bi-Py** dimers show a global increase in



**Figure 10.10.** Molecular models of poly guanine tracts. a) Comparison of A-form and B-form DNA with the sequence 5'-GATGGGGTAC-3'. The A-form DNA possesses a significantly wider, shallow and less curved minor groove. b) Comparison of the 5'-GGGG-3' and 5'-GAGG-3' sequences. Protons on the exocyclic amine of the guanine base are shown in yellow. Substitution of guanine for adenine removes the steric bulk of the exocyclic amine and replaces it with a proton.

affinity (10 to 50-fold) for the 5'-GGGG-3' target sequence in comparison to compound **1**. However, oligomers **4** and **6** are not able to clearly distinguish between the 5'-GGGG-3' match site and the 5'-GGGA-3' or 5'-GAGG-3' mismatch sites. Oligomers **3** and **5** are able to discriminate between the G-tetrad match sequence and the designed mismatches with 3- to 5-fold selectivity, respectively.

Interestingly, while oligomers **3-6** all display different curvatures, **5** is closest to mimicking the position of the N(3) atoms of the parent compound **1**, and may be more complementary to the A-form 5'-GGGG-3' helix. Compound **5** is

also the most successful of the series with respect to an increase in binding affinity, while maintaining reasonable specificity for the 5'-GGGG-3' DNA sequence.

In general, the global increase in affinity for these novel oligomers is not altogether unexpected. In contrast to the 5-membered heterocyclic carboxamides, the 6-5 fused benzimidazole analogues have a larger hydrophobic surface, likely promoting their placement in the DNA minor groove.<sup>60</sup> Aromatic stacking and van der Waals interactions are also contributing factors.<sup>44, 61</sup> Furthermore, the benzimidazole derivatives are a more rigid structure with a lower degree of rotational freedom. Such pre-organization may decrease the entropic cost of DNA complexation. Finally, the overall geometry of oligomers incorporated with benzimidazole derivatives is less curved, possibly allowing a more complementary fit with the DNA helix.

With respect to DNA sequence specificity, these oligomers show modest selectivity for their designed 5'-GGGG-3' match sites over the 5'-GGGA-3' and 5'-GAGG-3' mismatch sites. One explanation for the lower specificity is due to the inherent weakness of the designed mismatch sites. Replacing the exocyclic amine of guanine with the C(2) hydrogen of adenine results in the loss of a single hydrogen bond between the oligomer:DNA complex (Figure 10.10). Based on the footprinting titrations, an approximate value for that hydrogen bond is less than 0.7 kcal M<sup>-1</sup>. It is unclear whether oligomer:DNA complexation is driven more by enthalpic or entropic factors.<sup>52, 53, 60</sup> Furthermore, by placing an **Im** or **Ip** ring system across from the adenine base, there are no steric penalties, such as those that would be seen by positioning a pyrrole (**Py**) or benzimidazole (**Bi**) ring system over a guanine base.<sup>15, 62</sup> Subsequently, these oligomers show excellent sequence specificity for the 5'-GGGG-3' site in the presence of mixed G,C sequences.

**Conclusions.**

While traditional **Py**, **Im**, and **Hp** containing polyamides have been successful at recognizing hundreds of pre-determined DNA sequences with high affinity and specificity, a host of target sequences such as the G-tetrad (5'-GGGG-3') have proven difficult to code for. Based on structural data, the 5'-GGGG-3' sequence is believed to be highly A-form, with a wide, shallow, and undercurved DNA minor groove. Such structural features may indicate why past attempts at targeting this sequence with high affinity was unsuccessful. To address the recognition of this difficult sequence, a series of novel oligomers believed to be more complementary to the A-form helix, containing the 6-5 fused benzimidazole (**Bi**) and imidazopyridine (**Ip**) heterocycles, were developed. These oligomers, with a significantly different geometry than polyamides composed of 5-membered heterocyclic carboxamides, demonstrated a substantial increase in affinity (10 to 100-fold) for the 5'-GGGG-3' sequence. The marked increase in affinity is attributed to a combination of the oligomer properties including, a larger hydrophobic surface, a high degree of ligand pre-organization, and a greater complementarity to the A-form DNA helix. Future work directed towards improving sequence specificity and examination of the nuclear trafficking ability is a priority and will be reported in due course.

**Acknowledgments.** We thank the National Institutes of Health for grant support, Caltech for a James Irvine Fellowship to R.M.D., and the Parsons Foundation for fellowships to M.A.M and E. J. F.

## Experimental.

**General.** N,N-dimethylformamide (DMF), N,N-diisopropylethylamine (DIEA), thiophenol (PhSH), N,N-dimethylaminopropylamine (Dp), triethylamine (TEA), nitrobenzene (NO<sub>2</sub>Ph), 2-formyl-N-methylimidazole, red fuming nitric acid, 1,3-dichloro-4-nitropyridine, 30% bromine in acetic acid, palladium acetate (Pd(OAc)<sub>2</sub>), and 10% palladium on carbon, 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 60F<sub>254</sub> for TLC and silica gel 60 (40 μm) for flash chromatography were from Merck. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [ $\gamma$ -<sup>32</sup>P]triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) was from Amersham Pharmacia. DNaseI (7500 units/mL, FPLC pure) was purchased from Roche. AmpliTaq DNA polymerase was from Perkin-Elmer and used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-

borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d*<sub>6</sub> or CDCl<sub>3</sub> with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the California Institute of Technology. Matrix-assisted, laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was conducted at the Mass Spectroscopy Laboratory at the California Institute of Technology.

**Heterocycle Synthesis.** Heterocyclic monomers Boc-Py-OBt (**20**), Boc-Im-OH (**21**) and dimer Boc-Py-Bi-OH (**25**) were synthesized as reported.<sup>48</sup> Im-Im-OH (**21**) and Boc-Im-Im-OH (**24**), dimers were prepared using solution-phase dimerization methodology previously described.<sup>27</sup>

*1-methyl-4-nitro-1H-imidazole-2-carbaldehyde* (NO<sub>2</sub>-Im-CHO) (**13**). A cooled flask (0°C) of 1-methyl-2-imidazole-carboxaldehyde (**12**) (8g, 72.6 mmol, Aldrich) was treated dropwise with a precooled (0 °C) solution of red fuming nitric acid (75 ml) in conc. H<sub>2</sub>SO<sub>4</sub>•SO<sub>3</sub> (30%) (75 ml). The mixture was warmed to room temperature and stirred for 12 hours open to the atmosphere. Next, the mixture was poured over ice, neutralized with solid sodium carbonate, extracted four times with dichloromethane, dried over anhydrous sodium sulfate, and concentrated in vacuo to give a brownish-yellow oil. The oil was recrystallized from iPrOH/Et<sub>2</sub>O or EtOH/Et<sub>2</sub>O to give 1-methyl-

4-nitro-1H-imidazole-2-carbaldehyde (**13**) as a tan crystalline solid (4.5 g, 40% Yield). TLC (1:1 EtOAc/Hex)  $R_f$  0.4;  $^1\text{H}$  NMR (300 MHz, DMSO- $d_6$ )  $\delta$  9.74 (s, 1H), 8.71 (s, 1H), 3.99 (s, 3H);  $^{13}\text{C}$  (75 MHz, DMSO- $d_6$ )  $\delta$  182.31, 146.14, 140.56, 127.33, 35.70; HR-MS (EI+): calcd. for  $\text{C}_5\text{H}_5\text{O}_3\text{N}_3$ : 155.033; found: 155.035.

*2-(1-Methyl-4-nitroimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester* ( $\text{NO}_2\text{-Im-Ip-OMe}$ ) (**15**). 1-Methyl-4-nitro-1H-imidazole-2-carbaldehyde (2.01 g, 13.0 mmol) (**13**) and methyl 5,6-diaminopyridine-2-carboxylate (**11**) (2.17 g, 13.0 mmol) suspended in 120 ml of nitrobenzene was heated to 140 °C for 48 hours open to the atmosphere. The reaction mixture was cooled to 23 °C and the precipitate collected by vacuum filtration. The solid was washed with diethyl ether and dried under high vacuum to provide 2-(1-methyl-4-nitroimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (**15**) (3.6 g, 92% Yield) as a powdery tan solid.  $^1\text{H}$  NMR (300 MHz, DMSO- $d_6$ )  $\delta$  14.02 (broad s, 1H), 8.03-8.06 (m, 2H), 4.27 (s, 3H), 3.91 (s, 3H);  $^{13}\text{C}$  (75 MHz, DMSO- $d_6$ )  $\delta$  166.01, 146.44, 142.95, 136.59, 135.80, 130.36, 127.35, 123.81, 121.17, 121.04, 52.85, 37.43; HR-MS (EI+): calcd. for  $\text{C}_{12}\text{H}_{10}\text{N}_6\text{O}_4$ : 302.076; found: 302.076.

*2-(4-amino-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester* ( $\text{H}_2\text{N-Im-Ip-OMe}$ ) (**17**). 2-(1-methyl-4-nitroimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (**15**) (3 g, 9.9 mmol) was dissolved in anhydrous DMF (150 ml) and the solution was degassed with Ar. After the addition of Pd/C (10 wt. %, 600 mg) the reaction mixture was purged three times with hydrogen and then left to stir at 23°C for 9 hours under a hydrogen balloon atmosphere. After filtering through a pad of Celite and washing with copious amounts EtOAc the filtrate was

concentrated in vacuo to give 2-{4-amino-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (**17**) without further purification (2.7 g, 100% Yield).  $^1\text{H}$  NMR (300 MHz, DMSO- $d_6$ )  $\delta$  ;  $^{13}\text{C}$  (75 MHz, DMSO- $d_6$ )  $\delta$  166.17, 149.03, 148.21, 141.40, 131.50, 120.12, 107.77; HR-MS (EI+): calcd. for  $\text{C}_{12}\text{H}_{12}\text{N}_6\text{O}_2$ : 272.102; found: 272.103.

*2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (Boc-Im-Ip-OMe) (**18**).* 2-{4-amino-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (**17**) (2.0 g, 7.3 mmol) dissolved in DMF (25 ml) was treated with Boc<sub>2</sub>O (5.3 g, 24.3 mmol), DIEA (5.2 ml), and DMAP (95 mg, 0.73 mmol). The reaction mixture was then heated to 80 °C for 72 hours, cooled to 23 °C, and flashed through a plug of silica gel eluting with EtOAc to give a mixture of mono- and di-boced (2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3-[(tert-Butoxy)carbonylamino]-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester) products which were carried on for saponification.

*2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic (Boc-Im-Ip-OH) (**19**).* 2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3-[(tert-Butoxy)carbonylamino]-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester dissolved in MeOH (10 ml) and NaOH (1 N, 25 ml) was heated to 50 °C for 4 hours. The reaction mixture was cooled to 0 °C and the pH adjusted slowly to pH = 4 with 1 N HCl. The reaction mixture was then extracted with ethyl acetate (four times), dried over anhydrous sodium sulfate, concentrated in vacuo, and dried under high vacuum to give 2-{4-[(tert-Butoxy)carbonylamino]-1-



methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic (**19**) (258 mg, 60% Yield) as a light yellow solid.  $^1\text{H}$  NMR (300 MHz, DMSO- $d_6$ )  $\delta$  13.19 (broad s, 2H), 9.53 (s, 1H), 8.00 (m, 2H), 7.36 (s, 1H), 4.15 (s, 3H), 1.47 (s, 9H);  $^{13}\text{C}$  (75 MHz, DMSO- $d_6$ )  $\delta$  166.61, 152.92, 147.81, 142.48, 138.38, 132.01, 128.91, 119.74, 113.50, 79.07, 35.35, 28.12; HR-MS (EI+): calcd. for  $\text{C}_{16}\text{H}_{18}\text{N}_6\text{O}_4$ : 358.139; found: 358.137.

*2-(1-Methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester* (Im-Ip-OMe (**14**)). 1-Methylimidazole-2-carbaldehyde (**12**) (214 mg, 1.85 mmol) and methyl 5,6-diaminopyridine-2-carboxylate (**11**) (310 mg, 1.85 mmol) suspended in 17 ml of nitrobenzene was heated to 140 °C for 24 hours open to the atmosphere. The reaction mixture was cooled to room temperature. To the reaction was added diethyl ether (1 mL), upon which a tan precipitate formed. The precipitate was collected by filtration and washed with cold diethyl ether. The material was dissolved in hot iso propanol, cooled to room temperature and re-precipitated with diethyl ether. The solid was washed with diethyl ether and dried under high vacuum to provide 2-(1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (**14**) (250 mg, 52% Yield) as a tan solid.  $^1\text{H}$  NMR (300 MHz, DMSO- $d_6$ )  $\delta$  7.99 (s, 2 H), 7.51 (d, 1H,  $J = 0.9$  Hz), 7.20 (s, 1H,  $J = 0.9$  Hz), 4.18 (s, 3H), 3.88 (s, 3H);  $^{13}\text{C}$  (75 MHz, DMSO- $d_6$ )  $\delta$  147.7, 142.0, 137.1, 131.8, 129.6, 126.9, 119.9, 109.3, 52.8, 35.8; HR-MS (EI+): calcd. for  $\text{C}_{12}\text{H}_{11}\text{N}_5\text{O}_2$ : 257.091; found: 257.092.

*2-(1-Methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid* (Im-Ip-OH). (**16**). 2-(1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (**14**) (250 mg, 0.97 mmol) dissolved in MeOH (2 ml) and KOH (4 N, 3 ml) was heated to 50 °C for 4 hours. TLC showed baseline product formation with some

residual non-polar impurities. The methanol was removed in vacuo and the aqueous layer washed with EtOAc (2 x 10 mL) to remove any starting material and trace impurities. The pH of the aqueous layer was then adjusted slowly to pH = 4 with 1 N HCl upon which time a cloudy beige precipitate formed. The mixture was placed in a falcon tube and the precipitate concentrated by centrifugation. The supernatant was decanted and the solid dried under high vacuum to give 2-(1-Methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid (**16**) (154 mg, 65% Yield) as a brown solid. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>) δ 7.99 (s, 2H), 7.51 (d, 1H, *J* = 0.9 Hz), 7.20 (d, 1H, *J* = 0.9 Hz), 4.18 (s, 3H); HR-MS (EI<sup>+</sup>): calcd. for C<sub>11</sub>H<sub>9</sub>N<sub>5</sub>O<sub>2</sub>: 243.075; found: 243.074.

**Oligomer Synthesis.** Oligomers were synthesized on solid support using Boc-β-PAM resin (0.59 meq/g). Stepwise elongation of the oligomers was done according to previously published protocols.<sup>48</sup> The synthesis of compound **1** has been previously reported.<sup>32</sup>

*Preparation of Base Resin R-β-BiPy-BiPy-γ-NHBoc (BR1):* To a manual solid phase synthesis vessel was added Boc-β-PAM resin (0.3 g). The resin was washed with DMF (15 mL) and allowed to swell for 15 minutes while shaking at room temperature. The resin was then washed with DCM (~30 mL), followed by 80% TFA in DCM (~30 mL) to deblock the Boc-group. The resin was then agitated at room temperature in 80% TFA/DCM for another 25 minutes to provide the deblocked resin bound amine (**R-β-NH<sub>2</sub>**). Following Boc-deprotection, the resin was washed with DCM and 10% DIEA in DMF to neutralize and prepare for coupling. Simultaneously, in a separate reaction

vessel, Boc-PyBi-OH (**23**) (189 mg, 531  $\mu\text{M}$ ), HBTU (191 mg, 504  $\mu\text{M}$ ), DIEA (137 mg, 185  $\mu\text{L}$ , 1.06 mM) and DMF (1.2 mL) was mixed and allowed to activate at room temperature for 25 minutes. This mixture was then added to the solid-phase synthesis vessel containing **R**- $\beta$ -NH<sub>2</sub>. Coupling was allowed to proceed at room temperature with agitation for 3-6 hours. Initial loading of the resin requires elongated coupling times. Following coupling, the resin was acylated by the addition of acetic anhydride to the mixture and shaking for 15 minutes. The addition of the next Boc-PyBi-OH (**25**) dimer was incorporated and deprotected as described above to provide the resin bound fragment (**R**- $\beta$ -BiPy-Bi-Py-NH<sub>2</sub>). To this fragment was added a preactivated mixture of Boc- $\gamma$ -OH (180 mg, 885  $\mu\text{M}$ ), HBTU (319 mg, 841  $\mu\text{M}$ ), DIEA (229 mg, 308  $\mu\text{L}$ , 1.77 mM). Coupling was allowed to proceed for 3 h at room temperature with agitation. The resin was then capped with acetic anhydride as described above to provide the base resin **R**- $\beta$ -BiPy-Bi-Py- $\gamma$ -NHBoc (**BR1**). **BR1** was then washed with DCM followed by MeOH and Et<sub>2</sub>O. The resin was then dried under high vacuum and stored for subsequent use.

*Im-Im-Im-Im- $\gamma$ -PyBi-Py-Bi- $\beta$ -Dp* (**2**): **BR1** (50 mg) was added to a manual solid-phase synthesis vessel. The resin was washed with DCM (~15 mL), followed by deprotection with 80% TFA in DCM. The resin was shaken at room temperature in the 80% TFA solution for 25 minutes. The resin was then drained, washed with DCM, and neutralized with 10% DIEA in DMF. A pre-activated mixture of Boc-Im-Im-OH (**24**) (54 mg, 148  $\mu\text{M}$ ), HBTU (53 mg, 140  $\mu\text{M}$ ), DIEA (38 mg, 52  $\mu\text{L}$ , 295  $\mu\text{M}$ ) and DMF (400  $\mu\text{L}$ ) was then added to the reaction vessel and coupling was allowed to proceed for three hours at room temperature, followed by capping with acetic anhydride as described for **BR1** to give **R**- $\beta$ -BiPy-BiPy- $\gamma$ -Im-Im-NHBoc. Following resin deprotection as

described above, Im-Im-OH (**23**) was activated as described for Boc-Im-Im-OH (**24**). Coupling of **23** to the resin was allowed to proceed overnight at room temperature to provide **R**- $\beta$ -BiPy-BiPy- $\gamma$ -Im-Im-Im-Im. The resin was treated with the cleavage protocol outlined below to provide *Im-Im-Im-Im- $\gamma$ -PyBi-Py-Bi- $\beta$ -Dp* **2** in 5% Yield: [MALDI-TOF] C<sub>58</sub>H<sub>66</sub>N<sub>23</sub>O<sub>8</sub> 1212.55 calcd. 1212.5 found. [M+H]<sup>+</sup>

*ImIp-Im-Im- $\gamma$ -PyBi-Py-Bi- $\beta$ -Dp* (**3**): **BR1** (50 mg) was added to a manual solid phase synthesis vessel and **R**- $\beta$ -BiPy-BiPy- $\gamma$ -Im-Im-NHBoc was prepared as described above for **2**. Following deprotection, washing and neutralization as described above, a pre activated mixture of ImIp-OH (**14**) (21.5 mg, 88.5  $\mu$ M), HBTU (32 mg, 84  $\mu$ M), DIEA (23 mg, 31 $\mu$ L, 177  $\mu$ M), DMF (400  $\mu$ L) was added to the vessel containing **R**- $\beta$ -BiPy-BiPy- $\gamma$ -Im-Im-NH<sub>2</sub>. Coupling was allowed to proceed overnight at room temperature to provide **R**- $\beta$ -BiPy-BiPy- $\gamma$ -Im-Im-Im-Im. The resin was treated with the cleavage protocol outlined below to provide *ImIp-Im-Im- $\gamma$ -PyBi-Py-Bi- $\beta$ -Dp* **3** in 2.2% Yield: [MALDI-TOF] C<sub>59</sub>H<sub>64</sub>N<sub>23</sub>O<sub>7</sub> 1206.54 calcd. 1206.60 found. [M+H]<sup>+</sup>

*Im-Im-ImIp- $\gamma$ -PyBi-PyBi- $\beta$ -Dp* (**4**): **BR1** (50 mg, 0.81 meq/g) was added to a manual solid phase synthesis vessel. The resin was treated with the cleavage protocol outlined below to provide *Im-Im-ImIp- $\gamma$ -PyBi-PyBi- $\beta$ -Dp* **4** in 3% Yield: MALDI-TOF-MS C<sub>59</sub>H<sub>64</sub>N<sub>23</sub>O<sub>7</sub> 1206.5 calcd. 1206.5 found. [M+H]<sup>+</sup>

*Im-ImIp-Im- $\gamma$ -PyBi-PyBi- $\beta$ -Dp* (**5**): **BR1** (70 mg, 0.59 meq/g) was added to a manual solid phase synthesis vessel. The resin was washed with DCM (~15 mL), followed by deprotection with 80% TFA in DCM. The resin was shaken at room temperature in the 80% TFA solution for 25 minutes. The resin was then drained, washed with DCM, neutralized with 50% DIEA in DCM, and washed with DMF. A pre-

activated mixture of Boc-Im-OH (**21**) (50 mg, 207  $\mu\text{mol}$ ), HBTU (79 mg, 208  $\mu\text{mol}$ ), DIEA (53 mg, 72  $\mu\text{L}$ , 413  $\mu\text{mol}$ ) and DMF (900  $\mu\text{L}$ ) was then added to the reaction vessel and coupling was allowed to proceed for 12 hours at room temperature, followed by capping with acetic anhydride as described for **BR1** to give **R**- $\beta$ -BiPy-BiPy- $\gamma$ -Im-NHBoc. Following resin deprotection as described above, Boc-ImIp-OH (**19**) was activated as described for Boc-Im-OH (**21**). Coupling of **19** to the resin was allowed to proceed overnight at room temperature to provide **R**- $\beta$ -BiPy-BiPy- $\gamma$ -Im-Im-NHBoc. Following resin deprotection as described above, Im-CCl<sub>3</sub> (*2-Trichloroacetyl-1-methylpyrrole*) (47 mg, 207  $\mu\text{mol}$ ) and DIEA (53 mg, 72  $\mu\text{L}$ , 413  $\mu\text{mol}$ ) were dissolved in NMP (900  $\mu\text{L}$ ) and added to the reaction vessel. Coupling of Im-CCl<sub>3</sub> to the resin was allowed to proceed overnight at 32 °C to provide **R**- $\beta$ -BiPy-BiPy- $\gamma$ -Im-Im-Im. The resin was treated with the cleavage protocol outlined below to provide *Im-ImIp-Im- $\gamma$ -PyBi-Py-Bi- $\beta$ -Dp* **5** in 3% Yield: MALDI-TOF-MS C<sub>59</sub>H<sub>64</sub>N<sub>23</sub>O<sub>7</sub> 1206.54 calcd. 1206.5 found. [M+H]<sup>+</sup>

*ImIp-ImIp- $\gamma$ -PyBi-PyBi- $\beta$ -Dp* (**6**): **BR1** (50 mg) was added to a manual solid phase synthesis vessel. The resin was treated with the cleavage protocol outlined below to provide *ImIp-ImIp- $\gamma$ -PyBi-PyBi- $\beta$ -Dp* **6** in 2% Yield: MALDI-TOF-MS C<sub>60</sub>H<sub>62</sub>N<sub>23</sub>O<sub>6</sub> 1200.52 calcd. 1200.5 found. [M+H]<sup>+</sup>

**Resin Cleavage Procedure:** A sample of resin (20-100 mg) was washed with DCM followed by the addition of dimethylaminopropylamine (Dp) (1 mL). The mixture was heated to 80 °C for 2 h with occasional agitation. The resin was then filtered and washed with 0.1% TFA in water (7 mL). The combined filtrate was collected and

subjected to purification by reverse phase preparatory HPLC using a Waters C18 column and 0.1% TFA/ACN solvent system. Appropriate fractions from the HPLC purification were checked for purity by analytical HPLC and characterized by MALDI-TOF spectroscopy. Pure fractions were then pooled, flash frozen using liquid nitrogen and lyophilized to a dry solid for later use.

**Footprinting Experiments.** Plasmids pEF16 was constructed using standard methods.<sup>32</sup> DNase I footprint titrations were performed according to standard protocols.<sup>1</sup>

**References:**

- [1] Trauger, J. W.; Dervan, P. B., *Methods in Enzymology* **2001**, 340, 450-466.
- [2] Arcamone, F.; Nicoletti, V.; Penco, S.; Orezzi, P.; Pirelli, A., *Nature* **1964**, 203, (494), 1064.
- [3] Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology* **2003**, 13, (3), 284-299.
- [4] Wemmer, D. E., *Annual Review of Biophysics and Biomolecular Structure* **2000**, 29, 439-461.
- [5] Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001**, 9, (9), 2215-2235.
- [6] Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E., *Journal of Molecular Biology* **1985**, 183, (4), 553-563.
- [7] 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.
- [8] Klevit, R. E.; Wemmer, D. E.; Reid, B. R., *Biochemistry* **1986**, 25, (11), 3296-3303.
- [9] Pelton, J. G.; Wemmer, D. E., *Proceedings of the National Academy of Sciences of the United States of America* **1989**, 86, (15), 5723-5727.
- [10] Wade, W. S.; Mrksich, M.; Dervan, P. B., *Journal of the American Chemical Society* **1992**, 114, (23), 8783-8794.
- [11] 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.
- [12] Wade, W. S.; Dervan, P. B., *Journal of the American Chemical Society* **1987**, 109, (5), 1574-1575.
- [13] Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature* **1996**, 382, (6591), 559-561.
- [14] White, S.; Baird, E. E.; Dervan, P. B., *Chemistry & Biology* **1997**, 4, (8), 569-578.
- [15] Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nature Structural Biology* **1998**, 5, (2), 104-109.

- [16] White, S.; Baird, E. E.; Dervan, P. B., *Biochemistry* **1996**, 35, (38), 12532-12537.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Nature* **1998**, 391, (6666), 468-471.
- [21] Dervan, P. B.; Burli, R. W., *Current Opinion in Chemical Biology* **1999**, 3, (6), 688-693.
- [22] Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **2003**, 100, (21), 12063-12068.
- [23] Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B., *Nucleic Acids Research* **2004**, 32, (9), 2802-2818.
- [24] Belitsky, J. M.; Leslie, S. J.; Arora, P. S.; Beerman, T. A.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2002**, 10, (10), 3313-3318.
- [25] Doss, R. M.; Marques, M. A.; Foister, S.; Dervan, P. B., *Chemistry & Biodiversity* **2004**, 1, (6), 886-899.
- [26] Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2003**, 11, (20), 4333-4340.
- [27] Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B., *Helvetica Chimica Acta* **2002**, 85, (12), 4485-4517.
- [28] Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001**, 9, (1), 7-17.
- [29] Briehn, C. A.; Weyermann, P.; Dervan, P. B., *Chemistry-a European Journal* **2003**, 9, (9), 2110-2122.
- [30] Renneberg, D.; Dervan, P. B., *Journal of the American Chemical Society* **2003**, 125, (19), 5707-5716.



- [31] Marques, M. A.; Doss, R. M.; Foister, S.; Dervan, P. B., *Journal of the American Chemical Society* **2004**, 126, (33), 10339-10349.
- [32] Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1997**, 119, (30), 6953-6961.
- [33] Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1996**, 118, (35), 8198-8206.
- [34] Turner, J. M.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1997**, 119, (33), 7636-7644.
- [35] Kelly, J. J.; Baird, E. E.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **1996**, 93, (14), 6981-6985.
- [36] McCall, M.; Brown, T.; Kennard, O., *Journal of Molecular Biology* **1985**, 183, (3), 385-396.
- [37] Haran, T. E.; Shakked, Z.; Wang, A. H. J.; Rich, A., *Journal of Biomolecular Structure & Dynamics* **1987**, 5, (2), 199-217.
- [38] Lu, X. J.; Olson, W. K., *Nucleic Acids Research* **2003**, 31, (17), 5108-5121.
- [39] Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E., *Proceedings of the National Academy of Sciences of the United States of America* **1993**, 90, (7), 2930-2934.
- [40] Mutterer, F.; Weis, C. D., *Helvetica Chimica Acta* **1976**, 59, (1), 229-235.
- [41] Magerlein, W.; Indolese, A. F.; Beller, M., *Angewandte Chemie-International Edition* **2001**, 40, (15), 2856-2859.
- [42] Schoenbe, A.; Bartolet, I.; Heck, R. F., *Journal of Organic Chemistry* **1974**, 39, (23), 3318-3326.
- [43] Lombardy, R. L.; Tanious, F. A.; Ramachandran, K.; Tidwell, R. R.; Wilson, W. D., *Journal of Medicinal Chemistry* **1996**, 39, (7), 1452-1462.
- [44] Ji, Y. H.; Bur, D.; Hasler, W.; Schmitt, V. R.; Dorn, A.; Bailly, C.; Waring, M. J.; Hochstrasser, R.; Leupin, W., *Bioorganic & Medicinal Chemistry* **2001**, 9, (11), 2905-2919.
- [45] Behrens, C.; Harrit, N.; Nielsen, P. E., *Bioconjugate Chemistry* **2001**, 12, (6), 1021-1027.
- [46] Yadagiri, B.; Lown, J. W., *Synthetic Communications* **1990**, 20, (7), 955-963.

- [47] Austin, M. W.; Blackbor, Jr; Ridd, J. H.; Smith, B. V., *Journal of the Chemical Society* **1965**, (FEB), 1051.
- [48] Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1996**, 118, (26), 6141-6146.
- [49] Qu, X. G.; Ren, J. S.; Riccelli, P. V.; Benight, A. S.; Chaires, J. B., *Biochemistry* **2003**, 42, (41), 11960-11967.
- [50] Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., *Journal of Molecular Biology* **2002**, 320, (1), 55-71.
- [51] Melander, C.; Herman, D. M.; Dervan, P. B., *Chemistry-a European Journal* **2000**, 6, (24), 4487-4497.
- [52] Chaires, J. B., *Biopolymers* **1997**, 44, (3), 201-215.
- [53] Chaires, J. B., *Current Opinion in Structural Biology* **1998**, 8, (3), 314-320.
- [54] Koo, H. S.; Wu, H. M.; Crothers, D. M., *Nature* **1986**, 320, (6062), 501-506.
- [55] Urbach, A. R.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **2001**, 98, (8), 4343-4348.
- [56] Ng, H. L.; Dickerson, R. E., *Nucleic Acids Research* **2002**, 30, (18), 4061-4067.
- [57] Chalikian, T. V.; Plum, G. E.; Sarvazyan, A. P.; Breslauer, K. J., *Biochemistry* **1994**, 33, (29), 8629-8640.
- [58] Mrksich, M.; Dervan, P. B., *Journal of the American Chemical Society* **1995**, 117, (12), 3325-3332.
- [59] Mrksich, M.; Parks, M. E.; Dervan, P. B., *Journal of the American Chemical Society* **1994**, 116, (18), 7983-7988.
- [60] Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B., *Journal of Molecular Biology* **1997**, 271, (2), 244-257.
- [61] McKay, S. L.; Haptonstall, B.; Gellman, S. H., *Journal of the American Chemical Society* **2001**, 123, (6), 1244-1245.
- [62] Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., *Science* **1994**, 266, (5185), 646-650.