Chapter 5

DNA Minor Groove Recognition by Multiple Thiophene/Pyrrole Pairs

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

(Doss, R. M.; Marques. M. A. and Dervan, P. B. "*DNA Minor Groove Recognition by 3-Methylthiophene/Pyrrole Pair" Chemistry & Biodiversity* **2004**, *1*, 886-899.)

Abstract.

Hairpin polyamides are synthetic oligomers that fold and bind to specific DNA sequences in a *programmable* manner. Internal side-by-side pairings of the aromatic amino acids N-methylpyrrole (**Py**), N-methylimidazole (**Im**), and N-methylhydroxypyrrole (**Hp**) confer the ability to distinguish between all four Watson-Crick base pairs in the minor groove of B-form DNA. In a broad search to expand the heterocycle repertoire we found that when 3-methylthiophene (**Tn**), which presents a sulfur atom to the minor groove, is paired with **Py**, it exhibits a modest 3-fold specificity for T•A > A•T presumably by shape selective recognition. In this study we explore the scope and limitations of this lead by incorporating multiple **Tn** residues within a single hairpin polyamide. It was found that hairpin polyamides containing more that one **Tn/Py** pair exhibit lowered affinities and specificities for their match sites. It appears that little deviation is permissible from the parent 5-membered ring N-methylpyrrole-carboxamide scaffold for DNA recognition.

Introduction.

Polyamides, a class of crescent-shaped oligomers inspired by the natural products netropsin and distamycin A, are able to bind a broad repertoire of DNA sequences with affinities similar to naturally occurring proteins.[2-6] In the first-generation design, polyamide specificity can be attributed to the side-by-side pairings of N-methylpyrrole (**Py**) and N-methylimidazole (**Im**) aromatic rings in the minor groove of DNA where an **Im**/**Py** pair targets G•C and a **Py**/**Py** pair targets both A•T and T•A.[7] With the addition of the N-methylhydroxypyrrole (**Hp**) ring it was shown that the **Hp**/**Py** pair distinguishes T•A from A•T.[8, 9] The **Hp** ring exhibits specificity for T through steric fit and specific hydrogen bonds. The bump presented by the exocyclic hydroxyl group of **Hp** docks comfortably in the asymmetric cleft opposite T in a T•A base pair rather than suffer a sterically unfavorable interaction opposite the larger purine ring of A. From x-ray crystal structure analysis, it appears that **Hp** forms two specific hydrogen bonds with the O2 carbonyl of T.[8, 9]

Extensive studies were carried out on polyamides containing **Hp** in order to more fully explore how the incorporation of **Hp** affects ligand affinity and specificity.[10-12] Towards this end, several 8-ring hairpin polyamides were synthesized and their binding profiles assessed. According to the established pairing rules, a polyamide with the sequence Im-Im-**Hp**-Py-γ-Im-**Py**-Py-Py-β-Dp (**1**) (γ = gamma amino butyric acid; β = beta alanine; Dp = dimethylaminopropylamide) would be expected to target the six-base-pair site 5'-tGG**T**Ca-3' while a polyamide with the sequence Im-Im-**Py**-Py-γ-Im-**Hp**-Py-Py-β-Dp (**2**) should target 5'-tGG**A**Ca-3' (Figure 5.1a). The parent compound, Im-Im-Py-Py-γ-Im-Py-**Py**-Py-β-Dp (**3**) should target both A•T and T•A (5'-tGG**W**Ca-3') with similar affinities. Hairpins **1-3** were tested within the sequence context 5'-tGG**X**Ca-3' (**X** = A,T,G,C) where all four Watson-Crick base pairs were varied under the third (in bold) polyamide residue. As expected **1** bound its match site 5'-tGG**T**Ca-3' with ~20-fold preference over its mismatch sequence 5'-tGG**A**Ca-3' while **2** bound its match site 5'-tGG**A**Ca-3' with ~11-fold preference over its mismatch sequence 5'-tGG**T**Ca-3' (Table 5.1). **3** bound both sites 5'-tGG**A**Ca-3' and 5'-tGG**T**Ca-3' with similar affinities.[10]

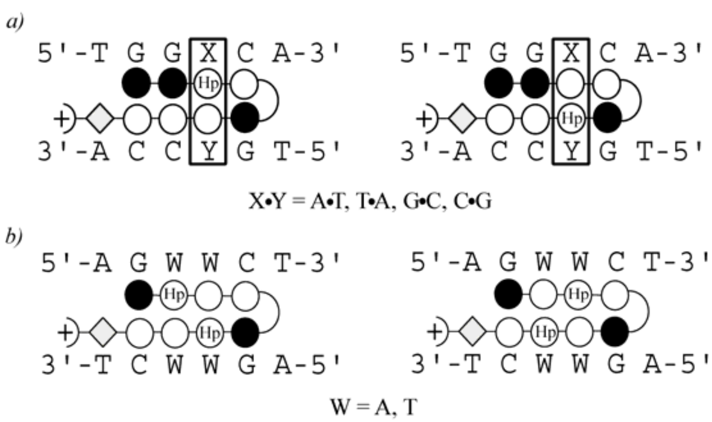


Figure 5.1. Dot models illustrating the examination of sequence selectivity of **Hp** against the four Watson-Crick base pairs (a) as well as hairpins containing multiple **Hp** residues against multiple A and T Watson-Crick base pairs (b). Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and hydroxypyrrole is indicated by a circle containing an **Hp**.

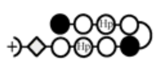
Table 5.1. *Hydroxypyrrole Hairpins:* K_a $[M^{-1}]^{a)}$)$^{b)}$)

Polyamide	A•T	T•A	G•C	C•G
	3.1 (±0.7)x 10^9	4.7 (±0.4)x 10^9	2.2 (±0.6)x 10^8	2.5 (±0.9)x 10^8
	8.1 (±1.9)x 10^7	1.6 (±0.3)x 10^9	5.5 (±1.5)x 10^7	7.9 (±2.1)x 10^7
	1.1 (±0.2)x 10^9	9.8 (±0.9)x 10^7	2.5 (±0.3)x 10^7	3.3 (±1.0)x 10^7

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

While results indicated that one could distinguish a single T•A base pair within a six-base-pair DNA site, a crucial next step was to explore how the incorporation of multiple **Hp** rings would be tolerated within a single hairpin polyamide. To address this question, polyamides Im-**Hp**-Py-Py-γ-Im-**Hp**-Py-Py-β-Dp (**4**) and Im-Py-**Hp**-Py-γ-Im-Py-**Hp**-Py-β-Dp (**5**) were designed to target their respective binding sites 5'-aG**TA**Ct-3' and 5'-aG**AT**Ct-3' (Figure 5.1b). We test whether all 4-ring pairings would code for a specific residue with each of the staggered **Hp** rings specifying for a T. As a control, polyamide Im-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**6**) was designed to bind both sequences with similar affinities. It was found that the specificities and affinities were significantly compromised by the incorporation of multiple **Hp** residues (Table 5.2).[12] Polyamide **4** bound its match site with a 50-fold reduction in affinity when compared to the parent compound **6**. Polyamide **5** bound with a lower affinity and preferred its double base-pair mismatch site over its designed match site.

Table 5.2. *Hairpins Containing Multiple Hyroxypyrrole Rings:* K_a [M^{-1}][a][b]

Polyamide	5'-aGTACt-3'	5'-aGAACt-3'	5'-aGATCt-3'
	$3.5 (\pm 0.7) x 10^{10}$	$4.7 (\pm 0.7) x 10^{9}$	$7.4 (\pm 1.5) x 10^{8}$
	$7.0 (\pm 1.8) x 10^{8}$	$\leq 1.0 x 10^{7}$	$\leq 1.0 x 10^{7}$
	$1.0 (\pm 0.2) x 10^{8}$	$2.6 (\pm 0.6) x 10^{7}$	$3.3 (\pm 0.7) x 10^{7}$

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

While **Hp** was a breakthrough ring for completing the four base pair code, it was clear that use of **Hp** would be limited for some sequence contexts. In addition, it was observed that oligomers containing **Hp** slowly degraded in the presence of acids or free radicals. This prompts us to examine the properties of other 5-membered heterocyclic amino acids as potential recognition elements for minor groove DNA recognition. Assuming that polyamide base pair specificity is derived, in part, from the functionality

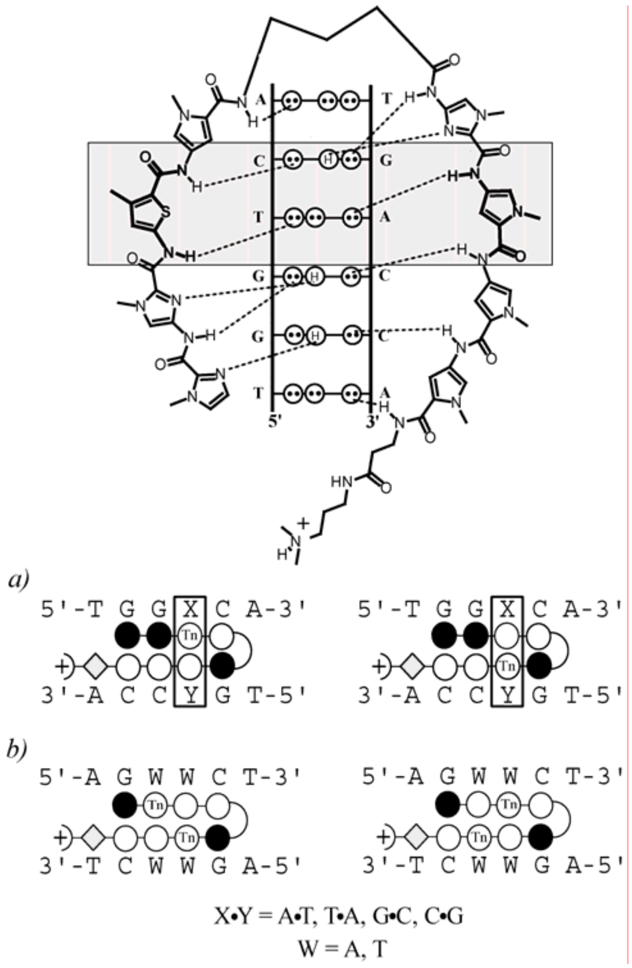


Figure 5.2 (Above) Hairpin polyamide and minor groove contacts for the match sequence. (Below) Dot models illustrating the examination of sequence selectivity of **Tn** against the four Watson-Crick base pairs (a) as well as hairpins containing multiple Tn residues against multiple A and T Watson-Crick base pairs (b). 3-methylthiophene is indicated by a circle containing a Tn. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and 3-methylthiophene is indicated by a circle containing a **Tn**.

presented to the minor groove floor by heterocycle ring pairs, we sought to explore new heterocycles for selective recognition.[13-17]

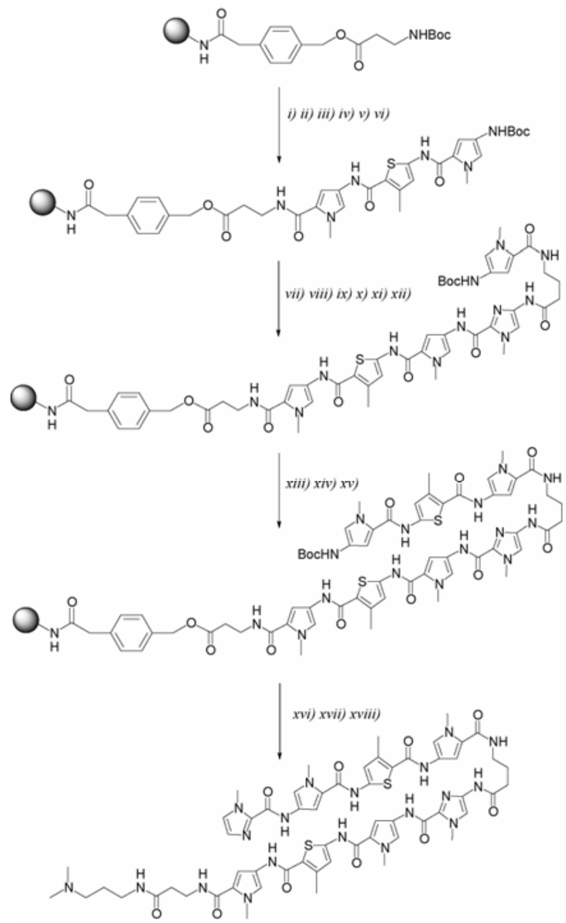
We have previously reported the sequence specificities (or lack thereof) of several novel rings systems when paired with (**Py**) at a single position within the hairpin polyamide sequence context Im-Im-**X**-Py-γ-Im-**Py**-Py-Py-β-Dp (**X** = 1-methylpyrazole (**Pz**), 1H-pyrrole (**Nh**), 5-methylthiazole (**Nt**), 4-methylthiazole (**Th**), 4-methylthiophene (**Tn**), thiophene (**Tp**), 3-hydroxythiophene (**Ht**), and furan (**Fr**). After an exhaustive study, it was found that 3-methylthiophene (**Tn**), exhibited modest specificity (~3 fold) for a T•A base pair when paired against **Py** and was able to maintain a high binding affinity at its match site of $K_a = 2.7 \times 10^9\ M^{-1}$.[14] **Tn** presents a large, polarizable sulfur atom to the minor groove and it is believed that its specificity for T is derived from the A•T base pair's ability to accommodate a large atom in the asymmetric cleft. The **Tn**/**Py** pairing was a potential step forward to replace **Hp** and we looked to explore the binding properties of hairpin polyamides containing more than one **Tn** residue (Figure 5.2).

While the selectivity of **Tn** for T•A > A•T was a modest, we were curious to see if there would be a multiplicity effect by targeting two T•A base pairs within a single hairpin binding site. Polyamides Im-**Tn**-Py-Py-γ-Im-**Tn**-Py-Py-β-Dp (**7**) and Im-Py-**Tn**-Py-γ-Im-Py-**Tn**-Py-β-Dp (**8**) were synthesized to test whether the overall base pair specificity would benefit from the incorporation of two specific **Tn** rings.

Results.

Monomer, Dimer, and Polyamide Synthesis. Hairpin polyamides were synthesized manually from Boc-β-PAM resin in a stepwise fashion using Boc-protected monomeric and dimeric amino acids according to solid-phase protocols (Figures 5.3). Boc-protected amino acid monomers and dimers for **Im**, **Py**, and **Tn-Im** were synthesized according to previously reported procedures.[13, 14] Synthesis of the **Tn-Py** amino acid dimer from the core amino ester (NH_2-Tn-OMe) is shown in Figure 5.4.

Tn-Py Dimer (**Tn-Py**). The hydrochloride salt of **9** (HCl•H_2N-Tn-OMe) was formed directly via cyclization reaction between **9** and amorphous sulfur.[14] The amine of **10** was Boc-protected using t-butyldicarbonate and DMAP to provide the Boc-protected



8

Figure 5.3. Solid phase synthetic scheme for Im-**Tn**-Py-Py-γ-Im-**Tn**-Py-Py-β-Dp starting from commercially available Boc-β-Pam resin: (i) 80% TFA/DCM, 0.4M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) Ac_2O, DIEA, DMF; (iv) 80% TFA/DCM, 0.4M PhSH; (v) Boc-Py-OBt, DIEA, DMF; (vi) Ac_2O, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) Boc-Im-Tn-OH, (HBTU, DIEA, DMF); (ix) Ac_2O, DIEA, DMF; (x) 80% TFA/DCM, 0.4M PhSH; (xi) Boc-γ-OH (HBTU, DIEA, DMF); (xii) Ac_2O, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) Boc-Py-OBt, DIEA, DMF; (xv) Ac_2O, DIEA, DMF; (xvi) 80% TFA/DCM, 0.4M PhSH; (xvii) Boc-Py-OBt, DIEA, DMF; (xviii) Ac_2O, DIEA, DMF; (xix) 80% TFA/DCM, 0.4M PhSH; (xx) Boc-Tn-OH, (HBTU, DIEA, DMF); (xxi) Ac_2O, DIEA, DMF; (xxii) 80% TFA/DCM, 0.4M PhSH; (xxiii) Im-$COCCl_3$ (DIEA, DMF); (xxiv) cleave from resin using (N,N-dimethylamino)propylamine, 85 ^oC.

ester **11** (Boc-Tn-OMe). The use of heat and the transacylation catalyst was necessary for the reaction to occur due to the poor reactivity of the thiophene aryl amine. Saponification of **11** was accomplished by heating in an aqueous solution of sodium hydroxide to provide **12** (Boc-**Tn**-OH). Alternatively, **10** was condensed with NO_2-**Py**-$COCCl_3$ in the presence of DMAP to provide the dimer **13** (NO_2-Py-**Tn**-OMe). The nitro group was reduced using a Parr apparatus (500 psi H_2) and Pd/C in a mixture of DMF and DIEA. Following reduction, t-butyl dicarbonate was added to the mixture to provide **14** (Boc-Py-**Tn**-OMe). **14** was saponified by heating in an aqueous solution of sodium hydroxide to provide **15** (Boc-Py-**Tn**-OH).

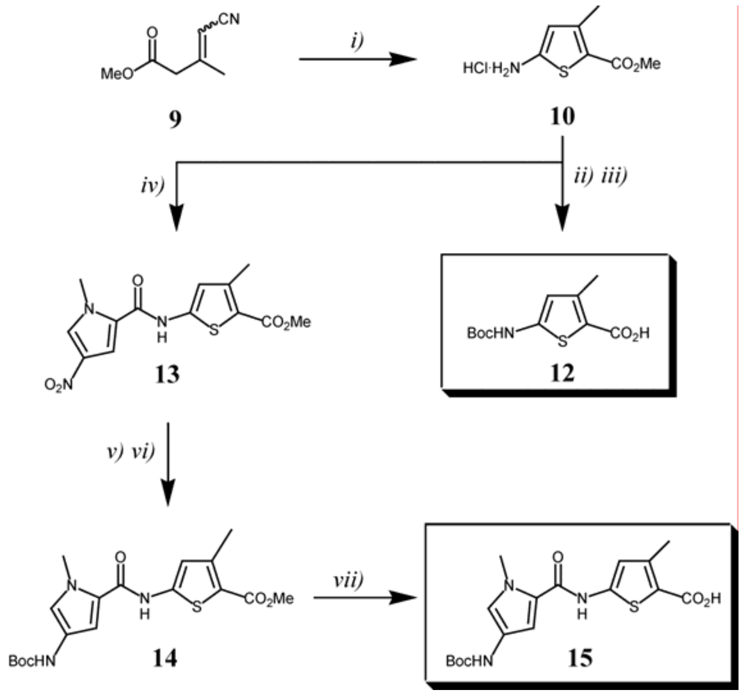


Figure 5.4. Synthesis of Boc-**Tn**-OH (**11**) and Boc-Py-Tn-OH (**14**). (i) S, Et_2NH, EtOH; (ii) $(Boc)_2O$, DMAP, DIEA, DMF; (iii) NaOH, MeOH; (iv) NO_2-Py-$COCCl_3$, DMAP, DIEA, EtOAc; (v) H_2 Pd/C, DIEA, DMF; (vi) $(Boc)_2O$; (vii) NaOH, MeOH.

DNA Affinity and Sequence Specificity in the Hairpin Motif. Quantitative DNase I footprinting titrations[18] were carried out for the following polyamides on the 285 bp PCR product of plasmids pDHN9 (polyamides **16** and **17**) and pDEH10 (polyamides **7** and **8**)[19]: Im-Im-**Tn**-Py-γ-Im-**Py**-Py-Py-β-Dp (**16**), Im-Im-**Py**-Py-γ-Im-**Tn**-Py-Py-β-Dp (**17**), Im-**Tn**-**Py**-Py-γ-Im-**Tn**-**Py**-Py-β-Dp (**7**), Im-**Py**-**Tn**-Py-γ-Im-**Py**-**Tn**-Py-β-Dp (**8**) (ring pairings of interest in bold). The DNA sequence specificity of each polyamide was determined by varying the DNA base pairs within the sequence context, 5'-tGG**X**Ca-3' (**X** = A, T, G, and C) for compounds **16** and **17**, and 5'-aG**WW**Ct-3' (**W** = A and T) for compounds **7** and **8** and comparing the relative affinities of the resulting complexes (Figure 5.5 & 5.6). The Watson-Crick base pairs were varied opposite the novel **Tn**/**Py** pairing in question, according to previously reported specificity studies on 8-ring hairpin polyamides.[14]

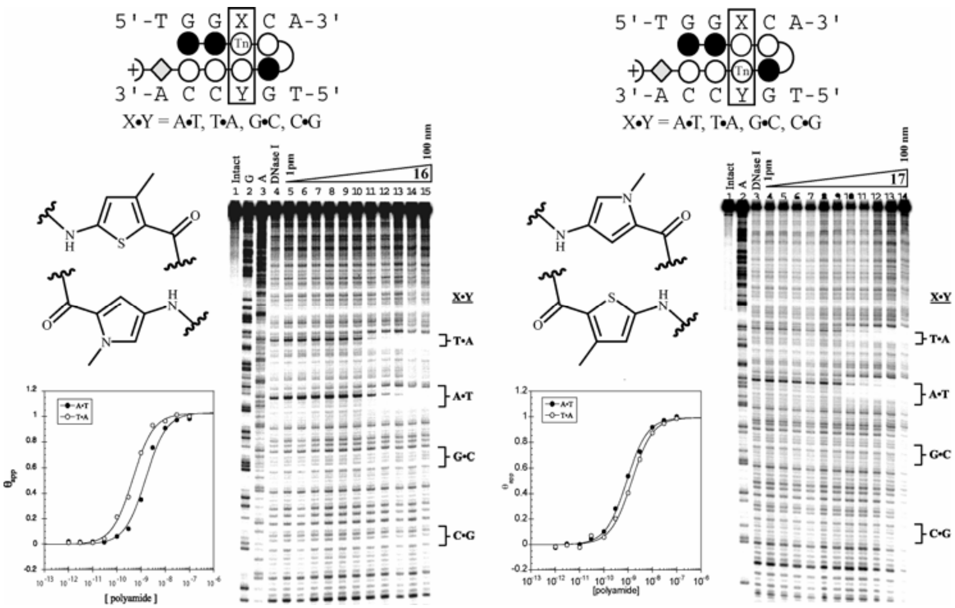


Figure 5.5. Quantitative DNase I footprint titration experiments for polyamides **16** and **17**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pDHN9: (A and B) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) chemical structure of the residue of interest; and (left bottom) binding isotherm 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; and Tn is indicated by a circle containing a **Tn**.

Hairpin **16** (**Tn/Py** pair) has been shown to bind with a high affinity for X = T, A ($K_a = 10^9$ M^{-1}), a 3-fold preference for T•A > A•T, and an 800-fold preference over the **X** = G, C sites (Figure 5.5, Table 5.3). The hairpin control **17**, which places the **Tn** ring across the polyamide, bound its match sequence with a reduced affinity ($K_a = 9.0 \times 10^8$ M^{-1}) and a lowered specificity of ~2 fold for T•A > A•T. It was found that both Im-**Tn**-**Py**-Py-γ-Im-**Tn**-**Py**-Py-β-Dp (**7**) and Im-**Py**-**Tn**-Py-γ-Im-**Py**-**Tn**-Py-β-Dp (**8**) bound their match sites with greatly reduced affinities ($K_a = 1.0 \times 10^8$ M^{-1} and $K_a = 4.5 \times 10^8$ M^{-1}

respectively) and exhibited non-specific binding at concentrations above 10nM (Figure 5.6, Table 5.4).

Table 5.3. *Thiophene Hairpins:* K_a [M^{-1}][a])[b])

Polyamide		A•T	T•A	G•C	C•G
	**3**	3.1 (±0.7)x 10^9	4.7 (±0.4)x 10^9	2.2 (±0.6)x 10^8	2.5 (±0.9)x 10^8
	**16**	8.0 (±0.4)x 10^8	2.7 (±0.2)x 10^9	≤ 10^6	≤ 10^6
	**17**	9.0 (±0.5)x 10^8	5.4 (±0.6)x 10^8	≤ 10^6	≤ 10^6

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

Discussion.

The search for a ring pair system that can successfully discriminate between the T•A and A•T base pairs has garnered much attention. After several extensive studies, we found that our best lead within the 5-member ring heterocycle family for sequence specificity lay in the modest 3-fold preference of the 3-methylthiophene ring for T over A. While attempts to selectively target multiple T•A base pairs with **Hp** were unsuccessful, we hoped that the **Tn** ring system would not suffer from the same reductions in affinity and specificity. 3-Hydroxypyrrole uses an exocyclic hydroxyl group as a means of shape-selective discrimination and although the 3-OH group can be tolerated by the relatively flexible T•A base pair, its size is slightly larger than optimal and may contribute to the reductions in affinity through clashes with the floor of the minor groove. In addition to steric issues, **Hp** containing hairpins may suffer an energetic penalty that stems from the hydration of 3-OH group.[20] In binding the minor groove of DNA, the polyamide is sequestered from the aqueous solvent and the

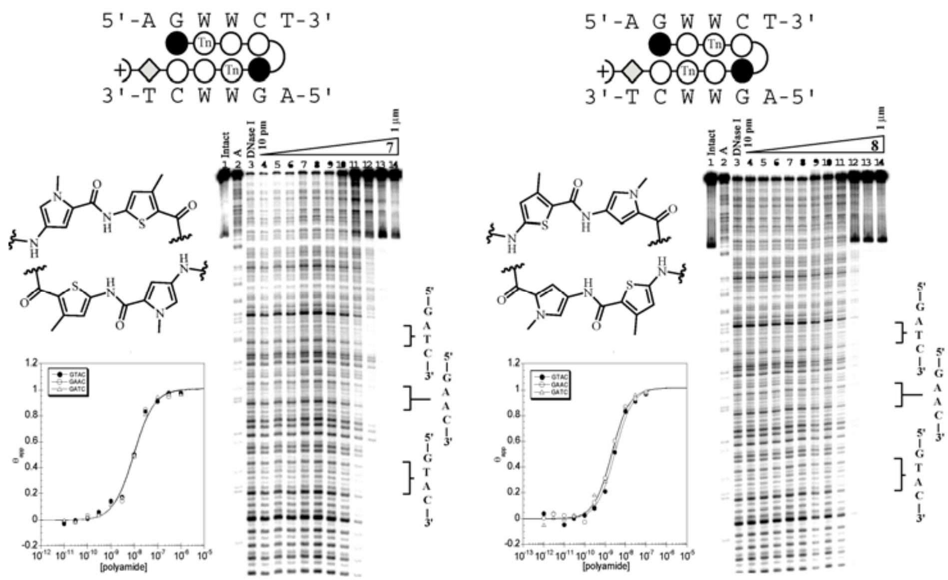
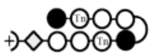
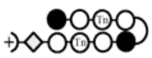


Figure 5.6. Quantitative DNase I footprint titration experiments for polyamides **7** and **8**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pDEH10: (A and B) lane 1, intact DNA; lane 2, G reaction; lane 3, DNase I standard; lanes 4-14, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) chemical structure of the residue of interest; and (left bottom) binding isotherm 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; and Tn is indicated by a circle containing a **Tn**.

differential hydration of the bound and unbound hairpins may contribute to the lowered affinities. 3-Methylthiophene, however, presents an endocyclic sulfur atom to the minor groove and solvation issues could be different.

We first examined whether the **Tn** ring's specificity for T•A would be conserved if the recognition element was moved from the top strand of the hairpin to the lower strand. Im-Im-**Py**-Py-γ-Im-**Tn**-Py-Py-β-Dp (**17**) was found to bind its match site 5'-atGG**A**Ca-3' with a moderate affinity and specificity of ~2-fold over its mismatch site. The reduction in affinity and specificity compared to the parent compound was

Table 5.4. *Hairpins Containing Multiple Thiophene Rings:* K_a [M^{-1}][a])[b])

Polyamide		5'-aGTACt-3'	5'-aGAACt-3'	5'-aGATCt-3'
	**6**	3.5 (± 0.7)x 10^{10}	4.7 (± 0.7)x 10^{9}	7.4 (± 1.5)x 10^{8}
	**7**	1.0 (± 0.5)x 10^{8}	1.0 (± 0.3)x 10^{8}	1.0 (± 0.4)x 10^{8}
	**8**	3.3 (± 0.9)x 10^{8}	4.7 (± 0.6)x 10^{8}	4.5 (± 0.7)x 10^{8}

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

anticipated from earlier studies. In translocating the **Tn** ring to the bottom strand, the match site for the hairpin was changed from 5'-at*GGT*Ca-3' to 5'-at*GGA*Ca-3'. It has been shown that certain DNA sequences, such as 5'-*GGA*-3' have lower affinities for hairpin polyamides, presumably due to altered B-form structure or lower intrinsic flexibility.

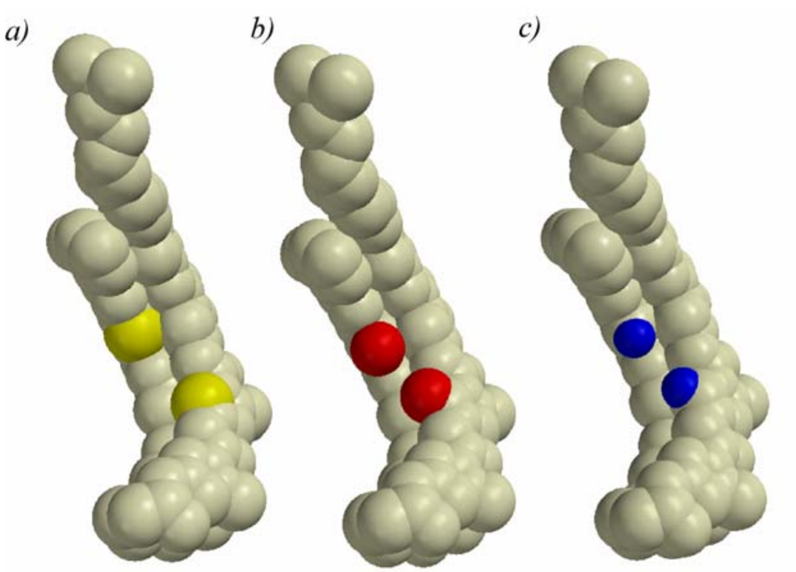


Figure 5.7. *Ab initio* models illustrate the differences in steric crowding that occur at the polyamide surface which is presented to the minor groove. In each model the atom which is varied is highlighted in a different color. A) Hairpin containing two, staggered **Tn** rings (sulfur in yellow). B) Hairpin containing two, staggered **Hp** rings (hydroxyl in red). C) Hairpin containing two, staggered Py rings (hydrogen in blue).

Incorporation of two **Tn** rings in polyamides **7** and **8** led to lower affinitites and an abolition of specificity at polyamide concentrations above 10nM. It may be that the minor groove is unable to accommodate two large sulfur atoms without disrupting the subtle steric interactions that confer the T selectivity of the **Tn** ring. *Ab initio* molecular modeling calculations using *Spartan Essential* software illustrate that there are significant steric differences between the **Py**, **Hp**, and **Tn** rings (Figure 5.7).[21] Binding affinities are consistent with the modeling trends as the **Tn** polyamide exhibits both the highest degree of steric crowding and the poorest binding affinity.

Conclusions.

Our search for novel recognition elements has again demonstrated that there is little room for deviation from the parent 5-membered ring N-methylpyrrole-carboxamide scaffold.[14] Although the **Tn/Py** pair can be used to selectively target a single T•A base pair, hairpin polyamides containing multiple **Tn/Py** pairs residues cannot be used to selectively target more than one T•A base pair. It should be noted that efforts to expand beyond N-methylpyrrole-carboxamide analogs to 6-5 fused bicycles (benzimidazole/hydroxybenzimidazole pairs) have shown promising levels of affinities and specificity for DNA.[22]

Experimental.

General. N,N-dimethylformamide (DMF), N,N-diisopropylethylamine (DIEA), thiophenol (PhSH), N,N-diethylamine, N,N-dimethylaminopropylamine (Dp), triethylamine (TEA), methyl 2-furoate, ketobutyric acid, methyl acetoacetate, cyanoacetic acid, trichloroacetyl chloride, pyrrole, sodium metal, methylthioglycolate, methyl-2-chloroacrylate, tin(II) chloride dihydrate, and thiourea were purchased from Aldrich.

Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam-resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), N,N-dimethylaminopyridine (DMAP), and Boc-β-alanine were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All other solvents were reagent grade from EM. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [γ-^{32}P]triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) and DNase I (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA polymerase was from Perkin-Elmer and used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d6* or $CDCl_3$ with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the California Institute of Technology. Matrix-assisted, laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was conducted at the Protein and Peptide Microanalytical Facility at the California Institute of Technology.

Heterocycle Synthesis. Boc-protected amino acid monomers and dimers for Boc-**Im-**OH, Boc-**Py-**OBt, and Boc-**Tn-Im**-OH were synthesized according to previously reported procedures.[13, 14]

Methyl 5- [(tert-butoxy)carbonylamino]-3-methylthiophene-2-carboxylate **(Boc-Tn-OMe**, **11**). A mixture of **10** (0.5 g, 2.40 mmol), $(Boc)_2O$ (1.58 g, 7.22 mmol), DIEA (622 mg, 839 μL, 4.81 mmol), DMAP (58 mg, 0.48 mmol) and DMF (5 mL) was stirred at 50 °C for 12 h. The solvent was removed *in vacuo* and the resulting brown residue subject to column chromatography (5:2 Hex/EtOAc). Rotoevaporation of the appropriate fractions provided a pale-yellow thin film, which when treated with hexanes and dried in vacuo gave **11** as an off-white solid (346 mg, 53% Yield). TLC (5:2 Hex/EtOAc) R_f 0.6; ^{1}H NMR (DMSO-d_6) 10.89 (s, 1H), 6.39 (s, 1H), 3.70 (s, 3H), 2.35 (s, 3H), 1.46 (s, 9H); ^{13}C (DMSO-d_6) 163.4, 152.8, 146.7, 145.5, 114.9, 114.7, 81.6, 52.0, 28.7, 16.6; EI-MS m/e 271.088 (M^+ calcd. for 271.088 $C_{12}H_{17}NO_4S$).

5- [(tert-butoxy)carbonylamino]-3-methylthiophene-2-carboxylate (**Boc-Tn-OH**, **12**). A mixture of **11** (300 mg, 1.1 mmol), MeOH (1 mL) and 1N NaOH (2 mL) was stirred at 60 °C for 4 h. The MeOH was removed in vacuo and the aqueous solution carefully adjusted to pH 2 with 1N HCl upon which time a milky suspension formed. The mixture was extracted with EtOAc (2 x 25 mL), dried over sodium sulfate. Removal of the organics in vacuo provided **12** as a tan powder (246 mg, 87% Yield). TLC (3:2

Hex/EtOAc, 10% AcOH) R_f 0.6; ^{1}H NMR (DMSO-d_6) 10.80 (s, 1H), 6.36 (s, 1H), 2.33 (s, 3H), 1.45 (s, 9H); ^{13}C (DMSO-d_6) 164.6, 152.7, 146.0, 144.6, 114.7, 81.4, 28.7, 16.5; EI-MS m/e 257.072 (M^+ calcd. for 257.072 $C_{11}H_{15}NO_4S$).

Methyl 3-methyl-5- [(1-methyl-4-nitropyrrol-2-yl)carbonylamino]thiophene-2-carboxylate (**NO_2-Py-Tn-OMe**, **13**). A mixture of **10** (1 g, 4.8 mmol), NO_2-Py-$COCCl_3$ (1.96 g, 7.21 mmol), DIEA (652 mg, 880 μL, 5.05 mmol), and DMAP (60 mg, 0.48 mmol) was stirred in EtOAc (15 mL) at 40 ^{o}C overnight. The mixture was allowed to cool to room temperature and sufficient hexanes were added to completely precipitate a pale-white solid. The precipitate was filtered, washed with cold EtOAc, and dried under vacuum to provide **13** (1.44 g, 93% Yield). TLC (5:2 Hex/EtOAc) R_f 0.55; ^{1}H NMR (DMSO-d_6) 11.73 (s, 1H), 8.27 (d, J = 1.8 Hz, 1H), 7.73 (d, J = 1.8 Hz, 1H), 3.95 (s, 3H), 3.73 (s, 3H), 2.42 (s, 3H); ^{13}C (DMSO-d_6) 163.5, 157.6, 144.8, 144.2, 134.6, 130.0, 125.0, 117.0, 116.5, 110.0, 52.1, 38.4, 16.3; EI-MS m/e 323.058 (M^+ calcd. for 323.058 $C_{13}H_{13}N_3O_5S$).

Methyl 5-({4- [(tert-butoxy)carbonylamino]-1-methylpyrrol-2-yl}carbonylamino)-3-methylthiophene-2-carboxylate (**Boc-Py-Tn-OMe**, **14**). A mixture of **13** (500 mg, 1.54 mmol), DIEA (400 mg, 536 μL, 3.08 mmol), Pd/C (50 mg) and DMF (6 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 $(Boc)_2O$ (500 mg, 2.28 mmol) was added. The mixture was then stirred for 8 h at 50 ^{o}C. The solvent was removed in vacuo, followed by column chromatography of the brown residue (5:2 Hex/EtOAc) to provide **13** as a pale-yellow film (205 mg, 34% Yield). TLC (5:2 Hex/EtOAc) R_f 0.37; ^{1}H NMR (DMSO-d_6) 11.36 (s, 1H), 9.20 (s, 1H), 7.05 (s, 1H), 7.00

(s, 1H), 6.70 (s, 1H), 3.81 (s, 3H), 3.71 (s, 3H), 2.40 (s, 3H), 1.43 (s, 9H); ^{13}C (DMSO-*d*$_6$) 162.9, 157.8, 152.6, 144.5, 143.9, 122.8, 120.6, 118.9, 115.3, 114.8, 105.0, 59.7, 31.0, 22.1, 20.7, 14.1; EI-MS m/e 393.136 (M$^+$ calcd. for 393.136 $C_{18}H_{23}N_3O_5S$).

5-({4- [(tert-butoxy)carbonylamino]-1-methylpyrrol-2-yl}carbonylamino)-3-methylthiophene-2-carboxylic acid (**Boc-Py-Tn-OH, 15**). A mixture of **14** (200 mg, 0.51 mmol), MeOH (1 mL) and 1N NaOH (2 mL) was stirred at 60 ^{o}C for 4 h. The MeOH was removed in vacuo and the aqueous solution carefully adjusted to pH 2 with 1N HCl upon which time a milky white precipitate formed. The mixture was extracted with EtOAc (2 x 25 mL), dried over sodium sulfate. Removal of the organics in vacuo provided **15** as a tan solid (160 mg, 83% Yield). TLC (3:2 Hex/EtOAc, 10% AcOH) R_f 0.6; ^{1}H NMR (DMSO-*d*$_6$) 11.28 (s, 1H), 9.20 (s, 1H), 7.04 (s, 1H), 6.99 (s, 1H), 6.68 (s, 1H), 3.81 (s, 3H), 2.38 (s, 3H), 1.43 (s, 9H); ^{13}C (DMSO-*d*$_6$) 164.1, 157.7, 152.6, 143.9, 143.1, 122.8, 120.7, 118.8, 116.5, 115.4, 105.0, 78.4, 36.3, 28.2, 15.8; EI-MS m/e 379.120 (M$^+$ calcd. for 379.120 $C_{17}H_{21}N_3O_5S$).

Hairpin 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.[13, 14]

*Im-**Tn**-Py-Py-γ-Im-**Tn**-Py-Py-β-Dp*: (Boc-Im-**Tn**-OH) (34 mg, 89 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-Py-β-Pam resin.

Coupling was allowed to proceed for 24 h at 40 °C, followed by capping with acetic anhydride 20% in DMF. After Boc-deprotection, Boc-γ-OH (18 mg, 89 μmol) was activated using HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Im-Tn-Py-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 40 °C, followed by capping with acetic anhydride 20% in DMF. After Boc-deprotection, the next two Py residues were incorporated as previously described. Boc-Tn-OH (23 mg, 89 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-Py-γ-Im-Tn-Py-Py-β-Pam resin. Coupling was allowed to proceed for 24 h at 40 °C followed by capping as described above. Boc-deprotection of the Boc-Tn-Py-Py-γ-Im-Tn-Py-Py-β-Pam resin was accomplished by shaking the resin in a 80% TFA in DCM mixture for 25 min at room temperature. The terminal **Im** residue was installed using Im-$COCCl_3$. Im-$COCCl_3$ (134 mg, 590 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (1 mL) were added to the H_2N-Tn-Py-Py-γ-Im-Tn-Py-Py-β-Pam resin and coupling was allowed to proceed for 48 h at 40 °C. The resin was then washed with DCM. Dp (1 mL) was added to the resin and the mixture was allowed to stand at 80 °C with occasional agitation for 2 h. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC to provide Im-**Tn**-Py-Py-γ-Im-**Tn**-Py-Py-β-Dp (**7**) (1.5 mg, 4.0% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1256.47 (M+H calcd. for 1256.50 $C_{58}H_{70}N_{19}O_{10}S$).

*Im-Py-**Tn**-Py-γ-Im-Py-**Tn**-Py-β-Dp*: (Boc-Py-**Tn**-OH) (34 mg, 89 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-β-Pam resin. Coupling was allowed to proceed for 24 h at 40 ^oC, followed by capping with acetic anhydride 20% in DMF. After Boc-deprotection, Boc-γ-Im-OH (29 mg, 89 μmol) was activated using HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Im-Py-Tn-Py-β-Pam resin. Coupling was allowed to proceed for 4 h at room temperature, followed by capping. After Boc-deprotection, the Py residue was incorporated as previously described (Ref). The next Boc-Py-Tn-OH dimer was incorporated as described above. After Boc-deprotection, the final **Im** residue was added using Im-$COCCl_3$. Im-$COCCl_3$ (134 mg, 590 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (1 mL) were added to the H_2N-Py-Tn-Py-γ-Im-Py-Tn-Py-β-Pam resin and coupling was allowed to proceed for 2 h at 40 ^oC. The resin was then washed with DCM. Dp (1 mL) was added to the resin and the mixture was allowed to stand at 80 ^oC with occasional agitation for 2 h. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC to provide Im-Py-**Tn**-Py-γ-Im-Py-**Tn**-Py-β-Dp (**8**) (1.9 mg, 5.1% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1256.50 (M+H calcd. for 1256.50 $C_{58}H_{70}N_{19}O_{10}S$).

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.

References:

[1] Trauger, J. W.; Dervan, P. B., *Methods in Enzymology* **2001,** 340, 450-466.

[2] Dervan, P. B.; Edelson, B. S., *Current Opinion in Structural Biology* **2003,** 13, (3), 284-299.

[3] Dervan, P. B.; Burli, R. W., *Current Opinion in Chemical Biology* **1999,** 3, (6), 688-693.

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

[5] Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature* **1996,** 382, (6591), 559-561.

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

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Melander, C.; Herman, D. M.; Dervan, P. B., *Chemistry-a European Journal* **2000,** 6, (24), 4487-4497.

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

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

[15] Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001,** 9, (1), 7-17.

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

[17] Ellervik, U.; Wang, C. C. C.; Dervan, P. B., *Journal of the American Chemical Society* **2000,** 122, (39), 9354-9360.

[18] Trauger, J. W., Dervan, P. B., *Methods Enzymol.* **2001,** 340, 450.

[19] Nguyen, D. H., Szewczyk, J. W., Baird, E. E., Dervan, P. B., *Bioorg. & Med. Chem.* **2001,** 9, 7.

[20] Wellenzohn, B., Loferer, M. J., Trieb, M., Rauch, C., Winger, R. H., Mayer, E., Liedl, K. R., *J. Am. Chem. Soc.* **2003,** 125, (4), 1088.

[21] SPARTAN ESSENTIAL Copyright © 1991-2001 by Wavefunction Inc.

[22] Renneberg, D., Dervan, P., *J. Am. Chem. Soc.* **2003,** 125, 5707.