Chapter 2

Completion of a Programmable DNA-Binding Small Molecule Library

Abstract

Hairpin pyrrole-imidazole (Py-Im) polyamides are programmable oligomers that bind the DNA minor groove in a sequence-specific manner with affinities comparable to those of natural DNA-binding proteins. These cell-permeable small molecules have been shown to enter the nuclei of live cells and downregulate endogenous gene expression. We complete here a library of 27 hairpin Py-Im polyamides which bind 7-base-pair sequences of the general form 5'-WWGNNNW-3' (where W = A or T, N = W, G, or C). Their equilibrium association constants (K_a) range from $K_a = 1 \times 10^8$ M^{-1} to 4×10^{10} M^{-1} with good sequence specificity. A table of binding affinities and sequence contexts for this completed 27-member library has been assembled for the benefit of the chemical biology community interested in molecular control of transcription.

2.1. Introduction

The biological applications of sequence-specific DNA-binding small molecules are a subject of intense research but still far from being routine.(1-11) Py-Im polyamides have been shown to influence a number of protein-DNA interactions, demonstrating both repression and activation of gene expression. Confocal microscopy experiments have confirmed that polyamide-fluorophore conjugates traffic unaided to the nuclei of living cells.(12-14) Since many diseases are attributed to aberrant gene expression, the regulation of transcriptional pathways with small molecules could have an important effect on human medicine.(15-17) For researchers interested in selectively targeting protein-DNA interfaces in promoters of specific genes with small molecules, access to well-characterized libraries of polyamides which bind a repertoire of different sequences with high affinity and specificity could enable development in this area.

Polyamides constructed from *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methylhydroxypyrrole (Hp) amino acids comprise a class of synthetic ligands that bind within the minor groove of DNA in a sequence-specific manner.(18,19) Inspired by the natural products netropsin and distamycin A, these programmed molecules recognize the Watson-Crick base pairs (bp) according to a series of pairing rules, where aromatic heterocycles paired in an antiparallel fashion are able to discriminate one Watson-Crick base pair from the other three combinations. The Py/Py pair recognizes A,T over C,G.(20) The Im/Py pair distinguishes G•C from C•G.(21,22) The exocyclic amine of guanine presents steric hindrance to the C3-H of Py, while the N3 of Im accommodates the amine and accepts one of its hydrogen bonds. The Hp/Py pair discriminates T•A over A•T due to the steric fit of the hydroxy group protruding into the minor groove, thus completing the pairing rules.(23,24) NMR and X-ray crystallography studies reveal that the crescent-shaped polyamide side-by-side dimer binds B-form DNA, a remarkable example of shape-selective recognition of the deep minor groove of DNA.(20,22,25,26)

Within the framework of the pairing rules, covalent linkages between two antiparallel

polyamide strands result in several possible structures, including the hairpin, cycle, H-pin, and U-pin binding motifs.(27-30) These linked structures show improved affinity and specificity when compared with the unlinked dimers. The eight-ring hairpin polyamide provides a good compromise between synthetic ease (linear vs. branched oligomers) and molecular recognition properties. In this binding motif, a γ-aminobutyric acid residue connects the carboxylic terminus of one strand to the amino terminus of the other.(27) The turn residue also serves as a DNA recognition element, as it has been shown to bind A•T and T•A base pairs with greater than 25-fold specificity over C•G and G•C, presumably for steric reasons.(31) Use of a chiral diaminobutyric acid turn residue increases the overall binding affinity of the molecule by 10-fold without a loss of sequence specificity.(32)

When hairpin polyamides are synthesized using solid phase methods on Boc-β-Ala-PAM resin and cleaved with 3-dimethylamino-1-propylamine, the product contains a β-alanine residue and a dimethylaminopropylamide tail at the C-terminus. Both of these elements are specific for W (where W = A or T) over G•C and C•G, again for steric reasons. In the hairpin motif, the β-alanine residue exhibits greater than 210-fold specificity for A•T and T•A base pairs over G•C and C•G.(31) The dimethylaminopropylamide tail shows a 20-fold preference for A•T and T•A over G•C and C•G.(31) Therefore, an eight-ring hairpin polyamide can bind seven base pairs with specificity for W over the turn, β-alanine residue, and tail (Figure 2.1).

We have excluded Hp in this study in favor of more stable Py residue. The electron-rich Hp ring degrades in the presence of strong acid and free radical impurities. Based on the paradigm of unsymmetrical ring pairings, more robust rings such as the hydroxybenzimidazole/Py pair can replace Hp for T•A recognition.(33) In our work, hairpin polyamide design typically includes an Im/Py pairing at the hairpin terminus in order to impart G•C specificity at this position. As a result, eight-ring pyrrole-imidazole hairpin polyamides can specifically bind 7-bp sequences of the general form 5'-WWGNNNW-3' (where N = W, G, or C). There are 27 possible permutations that fall within these
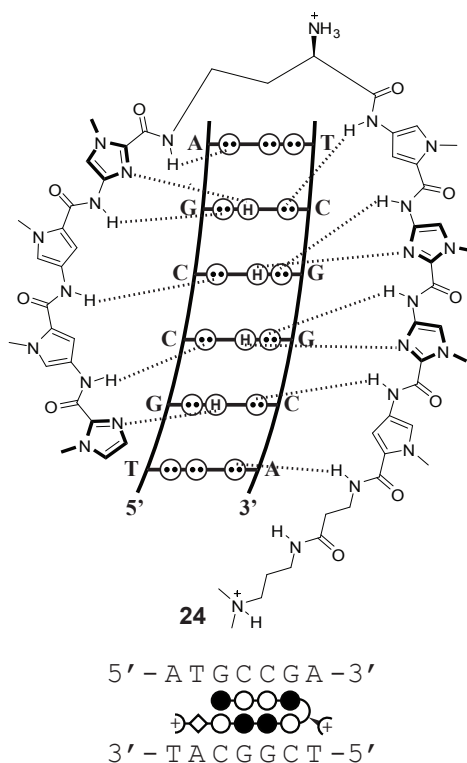
Figure 2.1. (Top) Model for the complex formed between hairpin polyamide **24** and its match DNA sequence. Circles with two dots represent the lone pairs of N(3) of purines and O(2) of pyrimidines. Circles containing an H represent the N(2) hydrogen of G. Hydrogen bonds are illustrated by dotted lines. (Bottom) Ball-and-stick binding model for the hairpin motif with the polyamide bound to its target DNA sequence. Imidazole and pyrrole are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the dimethylaminopropylamide tail is shown as a half-circle with a plus; and the chiral diaminobutyric acid turn residue is shown as a semicircle linked to a half-circle with a plus connecting the two subunits.

guidelines. Our group has published the energetics of 11 hairpin polyamides binding 11 distinct cognate sequences. Over half of the 27 sequences remain unreported. Searching our theses and notebooks confirmed 7 additional previously unpublished characterizations. Our studies during the past decade left only 9 out of 27 sequences unexamined; these are: 5'-WWGCGWW-3', 5'-WWGCCWW-3', 5'-WWGCWCW-3', 5'-WWGCCCW-3', 5'-WWGWCGW-3', 5'-WWGWGCW-3', 5'-WWGCGGW-3', 5'-WWGCCGW-3', and 5'-WWGWGGW-3'. To complete the library, we have synthesized nine hairpin polyamides designed to target these remaining sequences and assayed their binding affinity and sequence specificity by DNase I footprint titration experiments. In this study, we complete

the table of polyamide-DNA binding affinities, indicating each of the 27 general DNA sequences and an eight-ring hairpin polyamide that sequence specifically binds that 7-bp site. We hope that this centralized source of previously unpublished data proves helpful for other research groups currently modulating protein-DNA interfaces with DNA-binding small molecules.

2.2. Results

2.2.1. Polyamide synthesis. Nine polyamides **12**, **13**, **14**, **16**, **17**, **19**, **21**, **24** and **27** were synthesized on Boc-β-Ala-PAM resin according to published manual solid-phase synthesis protocols.(34) After cleavage with 3-dimethylamino-1-propylamine and reverse-phase HPLC purification, polyamides were characterized by analytical HPLC, UV-visible spectroscopy, and matrix-assisted laser desorption ionization/time of flight mass spectrometry.
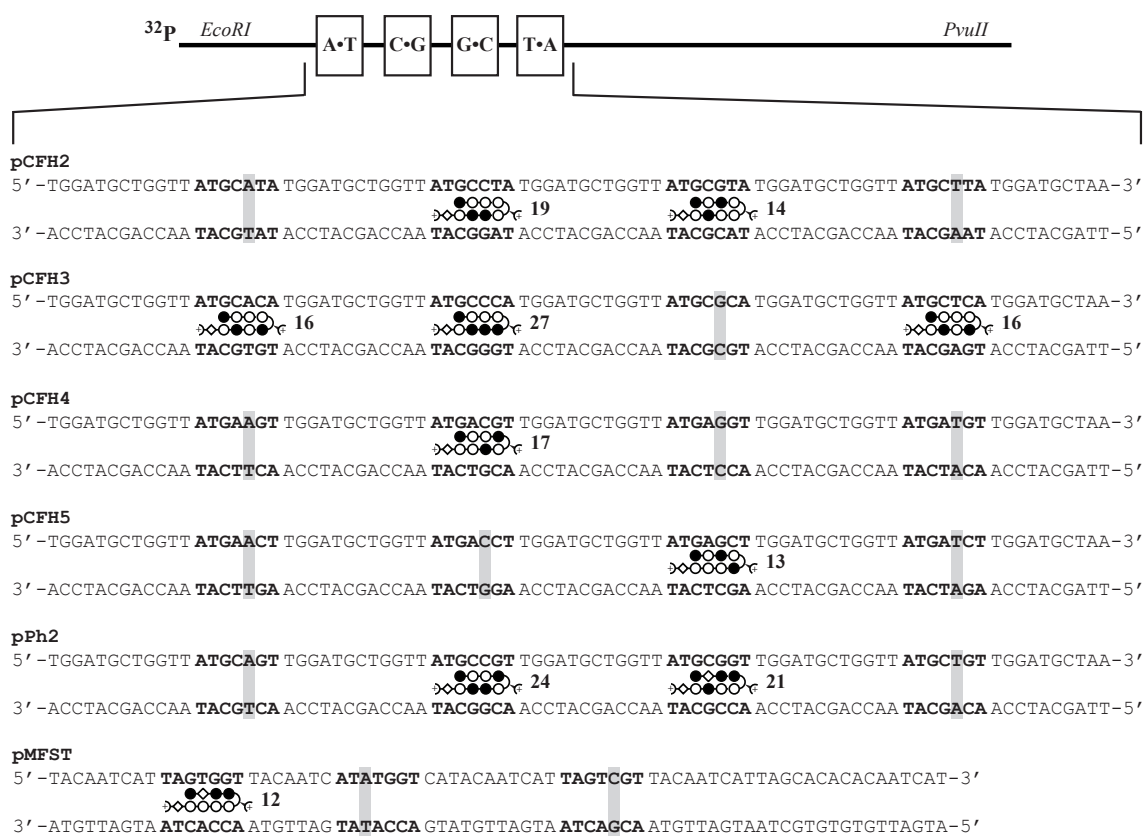

Figure 2.2. Plasmid design for pCFH2, pCFH3, pCFH4, pCFH5, pPh2, and pMFST, indicating the designed match and mismatch sites for hairpin polyamides **12**, **13**, **14**, **16**, **17**, **19**, **21**, **24**, and **27**. Imidazole and pyrrole are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the dimethylaminopropylamide tail is shown as a half-circle with a plus; and the chiral diaminobutyric acid turn residue is shown as a semicircle linked to a half-circle with a plus connecting the two subunits.

2.2.2. DNA binding energetics. Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0, 22°C) were performed on the 295 bp, 5'-^{32}P-end-labeled PCR product of plasmids pCFH2,
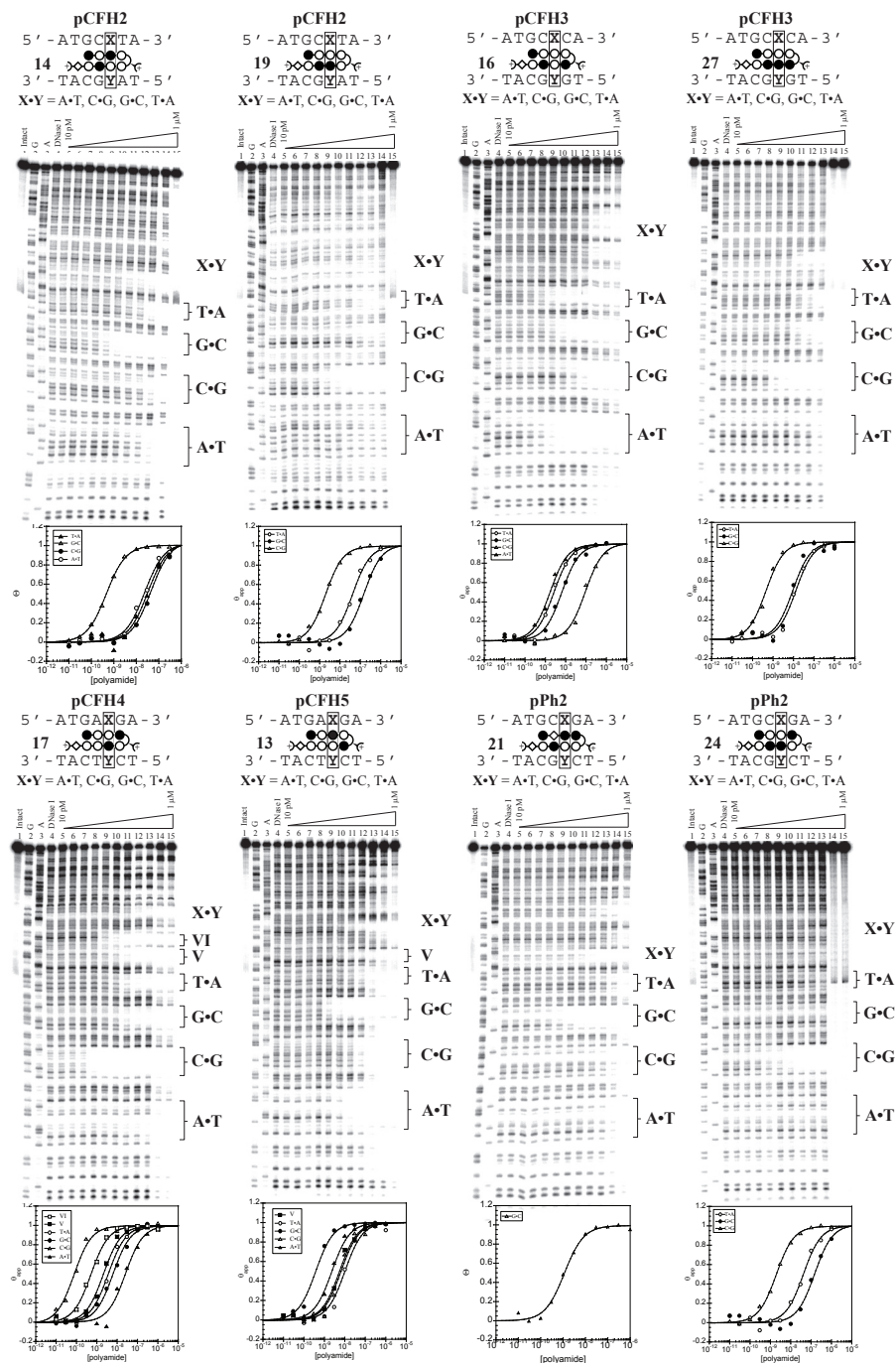
Figure 2.3. Quantitative DNase I footprint titration experiments for polyamides **13**, **14**, **16**, **17**, **19**, **21**, **24**, and **27** on the 295 bp, 5'-³²P-end-labeled PCR product of plasmids pCFH2, pCFH3, pCFH4, pCFH5, and pPh2. Lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase standard; lanes 5-15, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM, 300 nM, 1 μM, respectively. Binding isotherms for the four designed sites are shown below each footprinting gel; θ_norm values were calculated according to published methods. A binding model for the hairpin motif is shown above each gel with the polyamide bound to its target DNA sequence. Imidazole and pyrrole are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the dimethylaminopropylamide tail is shown as a half-circle with a plus; and the chiral diaminobutyric acid turn residue is shown as a semicircle linked to a half-circle with a plus connecting the two subunits.

pCFH3, pCFH4, pCFH5, pPh2, and pMFST.(35) Each of the nine polyamides was assayed on a plasmid containing its 7-bp match site according to the pairing rules, as well as three formal mismatch binding sites (Table 2.1 and Figure 2.2). The energetics of polyamide binding in the minor groove of DNA can be calculated from the Hill equation isotherms following DNase I cleavage and gel separation of the fragments. The equilibrium association constants (K_a) determined in this way provide a quantitative measure of polyamide affinity at a given DNA binding site. Comparing these constants across the four potential binding sites allows a relative measure of specificity for each base pair at the targeted position.
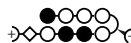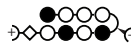
pCFH2 (5'→3')	ATGC**A**TA	ATGC**C**TA	ATGC**G**TA	ATGC**T**TA
14	$7.5 (\pm5.0)\times10^7$	$3.0 (\pm1.0)\times10^7$	**$2.1 (\pm0.2)\times10^9$**	$4.6 (\pm2.6)\times10^7$
19	$<1\times10^7$	**$7.0 (\pm2.2)\times10^8$**	$1.4 (\pm0.5)\times10^7$	$3.6 (\pm1.3)\times10^7$

pCFH3 (5'→3')	ATGC**A**CA	ATGC**C**CA	ATGC**G**CA	ATGC**T**CA
16	**$7.0 (\pm1.3)\times10^9$**	$2.5 (\pm1.3)\times10^8$	$2.5 (\pm1.0)\times10^9$	**$9.3 (\pm2.5)\times10^9$**
27	$<1\times10^7$	**$1.8 (\pm0.3)\times10^9$**	$9.8 (\pm1.3)\times10^7$	$7.1 (\pm0.6)\times10^7$

pCFH4 (5'→3')	ATGA**A**GA	ATGA**C**GA	ATGA**G**GA	ATGA**T**GA
17	$5.1 (\pm3.3)\times10^7$	**$1.3 (\pm0.1)\times10^{10}$**	$1.8 (\pm0.7)\times10^8$	$2.9 (\pm1.0)\times10^8$

pCFH5 (5'→3')	ATGA**A**CT	ATGA**C**CT	ATGA**G**CT	ATGA**T**CT
13	$5.3 (\pm1.4)\times10^8$	$9.5 (\pm4.4)\times10^7$	**$2.2 (\pm0.4)\times10^9$**	$8.2 (\pm4.1)\times10^7$

pPh2 (5'→3')	ATGC**A**GA	ATGC**C**GA	ATGC**G**GA	ATGC**T**GA
21	$<3.3\times10^7$	$<3.3\times10^7$	**$9.3 (\pm0.5)\times10^8$**	$<3.3\times10^7$
24	$<1\times10^7$	**$2.4 (\pm0.5)\times10^9$**	$1.1 (\pm0.3)\times10^7$	$2.2 (\pm1.4)\times10^7$

pMFST (5'→3')	TAGTGGT	ATATGGT	TAGTCGT
12	**$4.3 (\pm0.3)\times10^{10}$**	$9.9 (\pm0.6)\times10^8$	$6.2 (\pm0.2)\times10^9$

Table 2.1. K_a (M^{-1}) values reported are the mean values from at least three DNase I footprint titration experiments. 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. Imidazole and pyrrole are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the dimethylaminopropylamide tail is shown as a half-circle with a plus; the achiral γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and the chiral diaminobutyric acid turn residue is shown as a semicircle linked to a half-circle with a plus connecting the two subunits.

The nine polyamides **12**, **13**, **14**, **16**, **17**, **19**, **21**, **24** and **27** bind their respective 7-bp match sites with equilibrium association constants that range from $K_a = 7\times10^8$ M^{-1} to 4×10^{10}

	General (5'→3')	Polyamide	K_a (M^{-1})	Sequence context	Ref
1	WWGWWWW		3×10^9	5'-TA**G**TATT-3'	23
2	WWGGWWW		5×10^8	5'-C**TGG**TTA-3'	42
3	WWGWGWW		4×10^9	5'-TA**G**T**G**AA-3'	43
4	WWGWWGW		9×10^9	5'-TA**G**TA**G**T-3'	44
5	WWGWWCW		3×10^{10}	5'-TA**G**TA**C**T-3'	23
6	WWGWCWW		2×10^9	5'-**G**A**G**T**C**TA-3'	c
7	WWGCWWW		5×10^9	5'-AT**GC**AAA-3'	45
8	WWGGGWW		3×10^8	5'-AA**GGG**AA-3'	46
9	WWGGWGW		1×10^{10}	5'-TA**GG**T**G**T-3'	51
10	WWGGWCW		1×10^{10}	5'-AT**GG**T**C**A-3'	24
11	WWGGCWW		4×10^8	5'-AA**GGC**AT-3'	47
12	WWGWGGW		4×10^{10}	5'-TA**G**T**GG**T-3'	b
13	WWGWGCW		2×10^9	5'-AT**G**A**GC**T-3'	b
14	WWGCGWW		2×10^9	5'-AT**GCG**TA-3'	b
15	WWGCWGW		2×10^9	5'-TA**GC**A**G**T-3'	48
16	WWGCWCW		9×10^9	5'-AT**GC**T**C**A-3'	b
17	WWGWCGW		1×10^{10}	5'-AT**G**A**CG**T-3'	b
18	WWGWCCW		2×10^9	5'-TA**G**A**CC**A-3'	49
19	WWGCCWW		7×10^8	5'-AT**GCC**TA-3'	b
20	WWGGGGW		2×10^8	5'-**G**A**GGGG**T-3'	c
21	WWGCGGW		9×10^8	5'-AT**GCGG**T-3'	b
22	WWGGCGW		2×10^8	5'-CA**GGCG**T-3'	c
23	WWGGGCW		1×10^8	5'-C**TGGGC**A-3'	c
24	WWGCCGW		2×10^9	5'-AT**GCCG**T-3'	b
25	WWGGCCW		9×10^9	5'-AT**GGCC**A-3'	50
26	WWGCGCW		3×10^9	5'-AT**GCGC**A-3'	37
27	WWGCCCW		1×10^9	5'-AT**GCCC**A-3'	b

Table 2.2. Equilibrium association constants K_a (M^{-1}). Values reported are the mean values from at least three DNase I footprint titration experiments. [a]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. Imidazole and pyrrole are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the dimethylaminopropylamide tail is shown as a half-circle with a plus; the achiral γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and the chiral diaminobutyric acid turn residue is shown as a semicircle linked to a half-circle with a plus connecting the two subunits. [b] This paper, Table 2.1. [c] This paper, Figure 2.4.

2.3. Discussion

The microstructure of DNA is sequence dependent, and each hairpin polyamide acts as a microcaliper for the shape (width, depth) of the minor groove of DNA. Hence, it is not surprising that the energetics of hairpin binding for match sequences varies over two orders of magnitude (10^8-10^{10} M^{-1}) and is DNA sequence dependent. We would assume for transcriptional factor inhibition experiments one would prefer the highest affinity binding molecules. In the hairpin binding motif, β-alanine has been found to be a good structural replacement for Py, as the β/Im pair is specific for C•G and β/Py and β/β both code for A,T.(36,37) Furthermore, when β replaces Py adjacent to Im, binding is generally improved, as this flexible residue is believed to allow the amino acid pairings to reset their register with the DNA base pairs.(37,38) A key example of the use of β to replace Py can be seen in polyamide **26**. ImPyImPy-γ-ImPyImPy-β-Dp binds its target sequence 5'-ATGCGCA-3' with a binding affinity of 3×10^7 M^{-1}. Replacement of the internal Py residues with β residues yields the polyamide Im-β-ImPy-γ-Im-β-ImPy-β-Dp, which has a binding affinity of 3×10^9 M^{-1}, a 100-fold improvement.(37)

The chiral substitution of the γ-aminobutyric acid turn residue has been shown to improve binding affinity and sequence specificity.(32) For polyamides **14**, **16**, **19**, and **24**, the chiral diaminobutyric acid turn residue increases equilibrium association constants for these polyamides by 7- to 50-fold over their achiral variants (Table 2.3). In the example of polyamide **21**, the achiral pyrrole-imidazole compound (ImPyImIm-γ-PyPyImPy-β-Dp) binds its target sequence with an affinity of less than 1×10^7 M^{-1}, while the chiral version with β-alanine (Im-β-ImIm-(R)^{H2N}γ-PyPyImPy-β-Dp) exhibits greater than 90-fold enhancement, binding with an affinity of 9×10^8 M^{-1}. The combination of β-alanine substitution when appropriate and incorporation of the chiral turn can yield significantly improved DNA binding.

#	General (5'→3')	Achiral	K_a (M^{-1})	Sequence context	Ref	Chiral/β	K_a (M^{-1})	Sequence context	Ref
1	WWGWWWW	⬡	3×10^9	5'-TAGTATT-3'	23		-	-	-
2	WWGGWWW	⬡	5×10^8	5'-CTGGTTA-3'	42		-	-	-
3	WWGWGWW	⬡	7×10^8	5'-TAGTGAA-3'	43	⬡	4×10^9	5'-TAGTGAA-3'	43
4	WWGWWGW		-	-	-	⬡	9×10^9	5'-TAGTAGT-3'	44
5	WWGWWCW	⬡	3×10^{10}	5'-TAGTACT-3'	23		-	-	-
6	WWGWCWW		-	-	-	⬡	2×10^9	5'-GAGTCTA-3'	c
7	WWGCWWW		-	-	-	⬡	5×10^9	5'-ATGCAAA-3'	45
8	WWGGGWW	⬡	3×10^8	5'-AAGGGAA-3'	46		-	-	-
9	WWGGWGW		-	-	-	⬡	1×10^{10}	5'-TAGGTGT-3'	51
10	WWGGWCW	⬡	1×10^{10}	5'-ATGGTCA-3'	24		-	-	-
11	WWGGCWW	⬡	4×10^8	5'-AAGGCAT-3'	47		-	-	-
12	WWGWGGW		-	-	-	⬡	4×10^{10}	5'-TAGTGGT-3'	b
13	WWGWGCW		-	-	-	⬡	2×10^9	5'-ATGAGCT-3'	b
14	WWGCGWW	⬡	9×10^7	5'-ATGCGTA-3'	d	⬡	2×10^9	5'-ATGCGTA-3'	b
15	WWGCWGW		-	-	-	⬡	2×10^9	5'-TAGCAGT-3'	48
16	WWGCWCW	⬡	6×10^8	5'-ATGCACA-3'	d	⬡	9×10^9	5'-ATGCTCA-3'	b
17	WWGWCGW		-	-	-	⬡	1×10^{10}	5'-ATGACGT-3'	b
18	WWGWCCW	⬡	2×10^9	5'-TAGACCA-3'	49		-	-	-
19	WWGCCWW	⬡	1×10^8	5'-ATGCCTA-3'	d	⬡	7×10^8	5'-ATGCCTA-3'	b
20	WWGGGGW	⬡	2×10^7	5'-ATGGGGA-3'	50	⬡	2×10^8	5'-GAGGGGT-3'	c
21	WWGCGGW	⬡	$<1\times10^7$	5'-ATGCGGT-3'	d	⬡	9×10^8	5'-ATGCGGT-3'	b
22	WWGGCGW		-	-	-	⬡	2×10^8	5'-CAGGCGT-3'	c
23	WWGGGCW		-	-	-	⬡	1×10^8	5'-CTGGGCA-3'	c
24	WWGCCGW	⬡	4×10^7	5'-ATGCCGT-3'	d	⬡	2×10^9	5'-ATGCCGT-3'	b
25	WWGGCCW	⬡	9×10^9	5'-ATGGCCA-3'	50		-	-	-
26	WWGCGCW	⬡	3×10^7	5'-ATGCGCA-3'	50	⬡	3×10^9	5'-ATGCGCA-3'	37
27	WWGCCCW	⬡	$<2\times10^8$	5'-TAGCCCA-3'	45	⬡	1×10^9	5'-ATGCCCA-3'	b

Table 2.3. Equilibrium association constants K_a (M^{-1}). [a] K_a (M^{-1}) values reported are the mean values from at least three DNase I footprint titration experiments. 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. Imidazole and pyrrole are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the dimethylaminopropylamide tail is shown as a half-circle with a plus; the achiral γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and the chiral diaminobutyric acid turn residue is shown as a semicircle linked to a half-circle with a plus connecting the two subunits. [b] This paper, Table 1. [c] This paper, Figure 2.4. [d] Previously unpublished data.

2.4. Conclusion

We report here nine new eight-ring hairpin polyamides and assay their DNA-binding properties. These characterizations mark the completion of a 27-member polyamide library that covers all 7-bp DNA sequences 5'-WWGNNNW-3' (N = W, G, or C), where each compound binds its respective sequence with $K_a \geq 1 \times 10^8$ M^{-1}. Typical fold change for match versus single base-pair mismatch sites is in the range of 4 to 100. Looking forward, one could imagine using a DNA microarray-based approach to interrogate the entire sequence-recognition profile of each member of the 27 hairpin polyamide library on every permutation of DNA sites seven base pairs in size (42). We hope that this compilation of polyamide-DNA binding affinities will serve as a resource for ongoing small molecule gene regulation projects.

2.5. Experimental

2.5.1 Materials. (*tert*-Butoxycarbonyl)-β-alanine-PAM resin (Boc-β-Ala-PAM resin, divinylbenzene 1%, 200-400 mesh, 0.81 meq/g loading), N^α-9-fluorenylmethoxycarbonyl-N^γ-*tert*-butoxycarbonyl-D-2,4-diaminobutyric acid (Fmoc-D-Dab(Boc)-OH), and *O*-(benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU) were purchased from Peptides International. 1-Methyl-2-pyrrolidinone (NMP), *N,N*-diisopropylethylamine (DIEA), and 3-dimethylamino-1-propylamine (Dp) were purchased from Aldrich. *N,N*-Dimethylformamide (DMF) was purchased from EMD Biosciences, dichloromethane (DCM) was purchased from Fisher Scientific, and trifluoroacetic acid (TFA) was purchased from Halocarbon. Oligonucleotide inserts were synthesized by Integrated DNA Technologies. Bam HI, Hind III, polynucleotide kinase, DNase I, rapid DNA ligation kit, PCR core kit, and glycogen were purchased from Roche. JM109 competent cells and Wizard Plus Midipreps DNA Purification System were purchased from Promega. pUC19 plasmid DNA, β-mercaptoethanol, bromophenol blue, and xylene cyanole were purchased from Sigma. Calf thymus DNA and ProbeQuant G-50 Micro
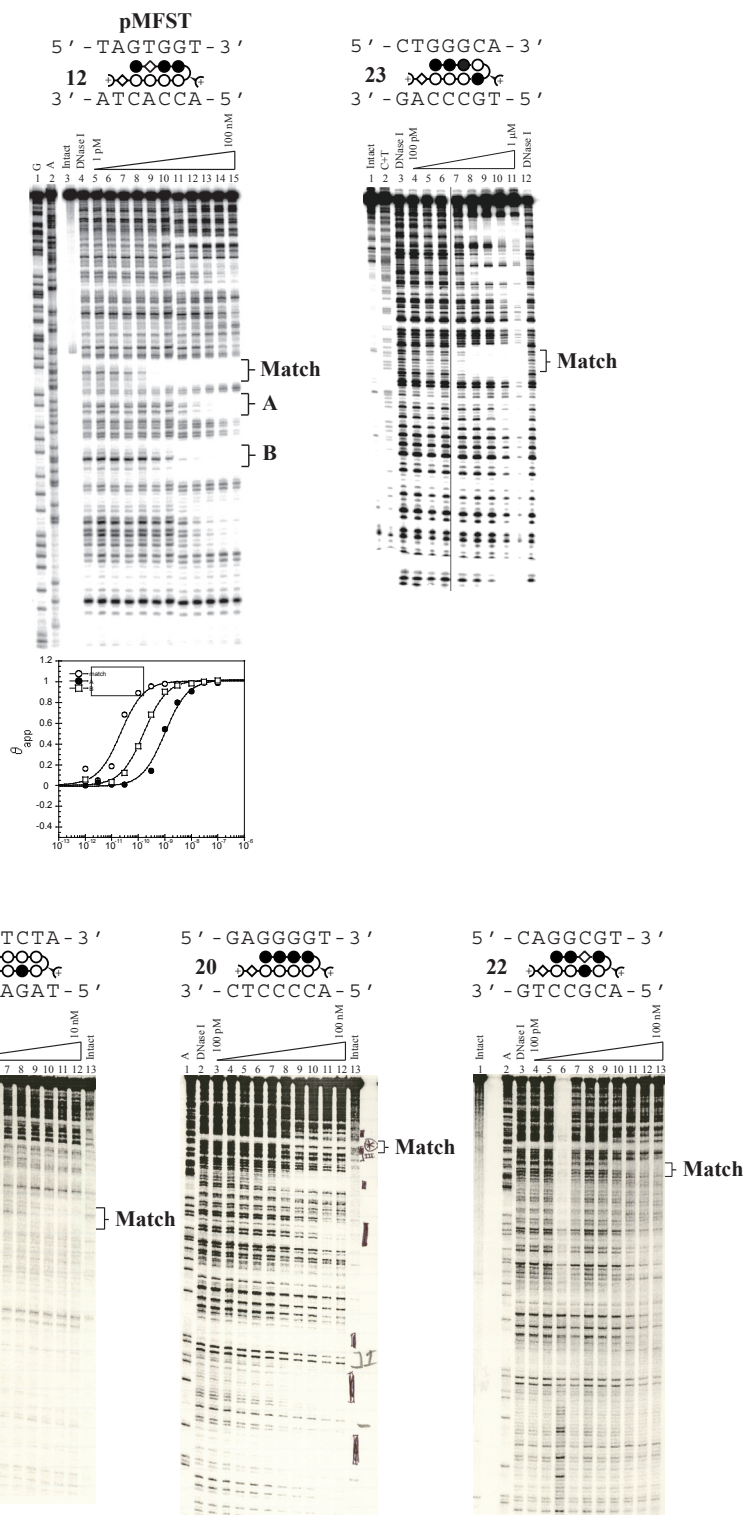
Figure 2.4. Quantitative DNase I footprint titration experiments for polyamides **6**, **12**, **20**, **22**, and **23**. A binding model for the hairpin motif is shown above each gel with the polyamide bound to its target DNA sequence. Imidazole and pyrrole are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the dimethylaminopropylamide tail is shown as a half-circle with a plus; and the chiral diaminobutyric acid turn residue is shown as a semicircle linked to a half-circle with a plus connecting the two subunits.

Columns were purchased from Amersham Biosciences. Adenosine 5'-triphosphate [γ-^{32}P], DL-dithiothreitol, and Tris were purchased from MP Biomedicals. 0.5 M EDTA, pH 8.0, phenol:chloroform:isoamyl alcohol (25:24:1 v/v), and formamide were purchased from Invitrogen. Potassium chloride, magnesium chloride, calcium chloride, and sodium chloride were purchased from Mallinckrodt. RNase-free water was purchased from USB. 8% Gene-PAGE PLUS, 7 M urea, denaturing acrylamide blend was purchased from Amresco. Tris borate EDTA was purchased from National Diagnostics. All reagents were used without further purification.

HPLC analysis was performed on a Beckman Gold system using a Phenomenex Gemini 4.6 × 250 mm, 5 μm 100 Å C_{18} reverse-phase column in 0.1% (w/v) TFA with acetonitrile as the eluent. Preparatory HPLC was carried out on a Beckman Gold system using either a Waters Delta-Pak 25 × 100 mm, 15 μm 300 Å C_{18} PrepPak Cartridge reverse-phase column or a Varian Dynamax 21.4 × 250 mm Microsorb 8 μm 300 Å C_8 reverse-phase column in 0.1% (w/v) TFA with acetonitrile as the eluent. UV spectra were measured on an Agilent Technologies 8453 UV-Vis ChemStation spectrophotometer. Matrix-assisted laser desorption ionization/time-of-flight mass spectrometry (MALDI-TOF-MS) was carried out on an Applied Biosystems Voyager DE-PRO. Storage phosphor autoradiography was performed on a Molecular Dynamics Typhoon 8600 phosphorimager. 18MΩ water was obtained from an AquaMAX Ultra water purification system, and all buffers were 0.2 μm filtered.

2.5.2. Polyamide synthesis. Polyamides were synthesized using pre-loaded Boc-β-Ala-PAM resin (50 mg, 0.81 meq/g) according to published manual solid-phase synthesis protocols.(34) The resin was cleaved with neat 3-dimethylamino-1-propylamine (1 mL) at 37°C with agitation for 16 h. Products were purified by preparatory reverse-phase HPLC and characterized by analytical HPLC, UV-visible spectroscopy, and MALDI-TOF mass spectrometry.

2.5.2.1. Im-β-ImIm-(*R*)^H2Nγ-PyPyPyPy-β-Dp (12). Polyamide **12** was isolated upon lyophilization of the appropriate fractions as a white powder (0.1 mg, 0.1% recovery). UV (H$_2$O) λ_{max} = 309 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1187.86 (1187.57 calcd for [M + H]$^+$ C$_{54}$H$_{71}$N$_{22}$O$_{10}$$^+$).

2.5.2.2. ImPyImPy-(*R*)^H2Nγ-ImPyPyPy-β-Dp (13). Polyamide **13** was isolated upon lyophilization of the appropriate fractions as a white powder (0.1 mg, 0.1% recovery). UV (H$_2$O) λ_{max} = 314 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1238.59 (1238.58 calcd for [M + H]$^+$ C$_{57}$H$_{72}$N$_{23}$O$_{10}$$^+$).

2.5.2.3. ImPyImPy-(*R*)^H2Nγ-PyPyImPy-β-Dp (14). Polyamide **14** was isolated upon lyophilization of the appropriate fractions as a white powder (0.1 mg, 0.1% recovery). UV (H$_2$O) λ_{max} = 315 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1238.58 (1238.58 calcd for [M + H]$^+$ C$_{57}$H$_{72}$N$_{23}$O$_{10}$$^+$).

2.5.2.4. ImPyPyPy-(*R*)^H2Nγ-ImPyImPy-β-Dp (16). Polyamide **16** was isolated upon lyophilization of the appropriate fractions as a white powder (0.4 mg, 1.5% recovery). UV (H$_2$O) λ_{max} = 315 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1238.49 (1238.58 calcd for [M + H]$^+$ C$_{57}$H$_{72}$N$_{23}$O$_{10}$$^+$).

2.5.2.5. ImPyPyIm-(*R*)^H2Nγ-PyImPyPy-β-Dp (17). Polyamide **17** was isolated upon lyophilization of the appropriate fractions as a white powder (1.8 mg, 7.1% recovery). UV (H$_2$O) λ_{max} = 316 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1238.63 (1238.58 calcd for [M + H]$^+$ C$_{57}$H$_{72}$N$_{23}$O$_{10}$$^+$).

2.5.2.6. ImPyPyPy-(*R*)^H2Nγ-PyImImPy-β-Dp (19). Polyamide **19** was isolated upon lyophilization of the appropriate fractions as a white powder (0.9 mg, 3.8% recovery). UV (H$_2$O) λ_{max} = 312 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1238.63 (1238.58 calcd for [M + H]$^+$ C$_{57}$H$_{72}$N$_{23}$O$_{10}$$^+$).

2.5.2.7. Im-β-ImIm-(*R*)^H2Nγ-PyPyImPy-β-Dp (21). Polyamide **21** was isolated upon lyophilization of the appropriate fractions as a white powder (0.3 mg, 0.2% recovery). UV (H$_2$O) λ_{max} = 309 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1188.67 (1188.57 calcd

for $[M + H]^+$ $C_{53}H_{70}N_{23}O_{10}{}^+$).

2.5.2.8. ImPyPyIm-(*R*)$^{\text{H2N}}$γ-PyImImPy-β-Dp (24). Polyamide **24** was isolated upon lyophilization of the appropriate fractions as a white powder (1.9 mg, 7.5% recovery). UV (H_2O) λ_{max} = 310 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1239.67 (1239.58 calcd for $[M + H]^+$ $C_{56}H_{71}N_{24}O_{10}{}^+$).

2.5.2.9. ImPyPyPy-(*R*)$^{\text{H2N}}$γ-ImImImPy-β-Dp (27). Polyamide **27** was isolated upon lyophilization of the appropriate fractions as a white powder (0.1 mg, 0.1% recovery). UV (H_2O) λ_{max} = 310 nm; MALDI-TOF-MS (monoisotopic) *m/z* 1239.60 (1239.58 calcd for $[M + H]^+$ $C_{56}H_{71}N_{24}O_{10}{}^+$).

2.5.3. DNase I footprinting experiments. Plasmids pCFH2, pCFH3, pCFH4, pCFH5, pPh2, and pMFST were constructed according to standard protocols for DNA manipulation.(39) PCR products (295 bp, 5'-^{32}P-end-labeled) were isolated and DNase I footprint titrations were performed according to standard protocols.(35) Chemical sequencing reactions were performed according to published methods.(40,41)

References

1. Gottesfeld, J.M., Neely, L., Trauger, J.W., Baird, E.E. and Dervan, P.B. (1997) Regulation of gene expression by small molecules. *Nature*, **387**, 202-205.

2. Dickinson, L.A., Gulizia, R.J., Trauger, J.W., Baird, E.E., Mosier, D.E., Gottesfeld, J.M. and Dervan, P.B. (1998) Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 12890-12895.

3. Mapp, A.K., Ansari, A.Z., Ptashne, M. and Dervan, P.B. (2000) Activation of gene expression by small molecule transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 3930-3935.

4. Janssen, S., Cuvier, O., Muller, M. and Laemmli, U.K. (2000) Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to Drosophila melanogaster. *Molecular Cell*, **6**, 1013-1024.

5. Maeshima, K., Janssen, S. and Laemmli, U.K. (2001) Specific targeting of insect and vertebrate telomeres with pyrrole and imidazole polyamides. *Embo Journal*, **20**, 3218-3228.

6. Gygi, M.P., Ferguson, M.D., Mefford, H.C., Lund, K.P., O'Day, C., Zhou, P., Friedman, C., van den Engh, G., Stolowitz, M.L. and Trask, B.J. (2002) Use of fluorescent sequence-specific polyamides to discriminate human chromosomes by microscopy and flow cytometry. *Nucleic Acids Research*, **30**, 2790-2799.

7. Crowley, K.S., Phillion, D.P., Woodard, S.S., Schweitzer, B.A., Singh, M., Shabany, H., Burnette, B., Hippenmeyer, P., Heitmeier, M. and Bashkin, J.K. (2003) Controlling the intracellular localization of fluorescent polyamide analogues in cultured cells. *Bioorganic & Medicinal Chemistry Letters*, **13**, 1565-1570.

8. Kodadek, T., Reddy, M.M., Olivos, H.J., Bachhawat-Sikder, K. and Alluri, P.G. (2004) Synthetic molecules as antibody replacements. *Accounts of Chemical Research*, **37**, 711-718.

9. Olenyuk, B.Z., Zhang, G.J., Klco, J.M., Nickols, N.G., Kaelin, W.G. and Dervan, P.B. (2004) Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 16768-16773.

10. Bando, T., Narita, A., Sasaki, S. and Sugiyama, H. (2005) Specific adenine alkylation by pyrrole-imidazole CBI conjugates. *Journal of the American Chemical Society*, **127**, 13890-13895.

11. Burnett, R., Melander, C., Puckett, J.W., Son, L.S., Wells, R.D., Dervan, P.B. and Gottesfeld, J.M. (2006) DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA·TTC repeats in Friedreich's ataxia. *Proceedings*

of the National Academy of Sciences of the United States of America, **103**, 11497-11502.

12.   Belitsky, J.M., Leslie, S.J., Arora, P.S., Beerman, T.A. and Dervan, P.B. (2002) Cellular uptake of N-methylpyrrole/N-methylimidazole polyamide-dye conjugates. *Bioorg Med Chem*, **10**, 3313-3318.

13.   Best, T.P., Edelson, B.S., Nickols, N.G. and Dervan, P.B. (2003) Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc Natl Acad Sci U S A*, **100**, 12063-12068.

14.   Edelson, B.S., Best, T.P., Olenyuk, B., Nickols, N.G., Doss, R.M., Foister, S., Heckel, A. and Dervan, P.B. (2004) Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res*, **32**, 2802-2818.

15.   Pandolfi, P.P. (2001) Transcription therapy for cancer. *Oncogene*, **20**, 3116-3127.

16.   Darnell, J.E. (2002) Transcription factors as targets for cancer therapy. *Nature Reviews Cancer*, **2**, 740-749.

17.   Arkin, M.R. and Wells, J.A. (2004) Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream. *Nature Reviews Drug Discovery*, **3**, 301-317.

18.   Dervan, P.B. (2001) Molecular recognition of DNA by small molecules. *Bioorganic & Medicinal Chemistry*, **9**, 2215-2235.

19.   Dervan, P.B. and Edelson, B.S. (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol*, **13**, 284-299.

20.   Pelton, J.G. and Wemmer, D.E. (1989) Structural Characterization of a 2-1 Distamycin A.D(Cgcaaattggc) Complex by Two-Dimensional Nmr. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 5723-5727.

21.   Wade, W.S., Mrksich, M. and Dervan, P.B. (1992) Design of Peptides That Bind in the Minor Groove of DNA at 5'-(a,T)G(a,T)C(a,T)-3' Sequences by a Dimeric Side-by-Side Motif. *Journal of the American Chemical Society*, **114**, 8783-8794.

22.   Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E. and Dervan, P.B. (1992) Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc Natl Acad Sci U S A*, **89**, 7586-7590.

23.   Trauger, J.W., Baird, E.E. and Dervan, P.B. (1996) Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature*, **382**, 559-561.

24. White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E. and Dervan, P.B. (1998) Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*, **391**, 468-471.

25. Kielkopf, C.L., Baird, E.E., Dervan, P.B. and Rees, D.C. (1998) Structural basis for G·C recognition in the DNA minor groove. *Nature Structural Biology*, **5**, 104-109.

26. Kielkopf, C.L., White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., Dervan, P.B. and Rees, D.C. (1998) A structural basis for recognition of A·T and T·A base pairs in the minor groove of B-DNA. *Science*, **282**, 111-115.

27. Mrksich, M., Parks, M.E. and Dervan, P.B. (1994) Hairpin peptide motif - a new class of oligopeptides for sequence-specific recognition in the minor-groove of double-helical DNA. *Journal of the American Chemical Society*, **116**, 7983-7988.

28. Herman, D.M., Turner, J.M., Baird, E.E. and Dervan, P.B. (1999) Cycle polyamide motif for recognition of the minor groove of DNA. *Journal of the American Chemical Society*, **121**, 1121-1129.

29. Greenberg, W.A., Baird, E.E. and Dervan, P.B. (1998) A comparison of H-pin and hairpin polyamide motifs for the recognition of the minor groove of DNA. *Chemistry-a European Journal*, **4**, 796-805.

30. Heckel, A. and Dervan, P.B. (2003) U-pin polyamide motif for recognition of the DNA minor groove. *Chemistry-a European Journal*, **9**, 3353-3366.

31. Swalley, S.E., Baird, E.E. and Dervan, P.B. (1999) Effects of gamma-turn and beta-tail amino acids on sequence-specific recognition of DNA by hairpin polyamides. *J Am Chem Soc*, **121**, 1113-1120.

32. Herman, D.M., Baird, E.E. and Dervan, P.B. (1998) Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides in the minor groove. *Journal of the American Chemical Society*, **120**, 1382-1391.

33. Renneberg, D. and Dervan, P.B. (2003) Imidazopyridine/pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *Journal of the American Chemical Society*, **125**, 5707-5716.

34. Baird, E.E. and Dervan, P.B. (1996) Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *Journal of the American Chemical Society*, **118**, 6141-6146.

35. Trauger, J.W. and Dervan, P.B. (2001) Footprinting methods for analysis of pyrrole-imidazole/DNA complexes. *Methods in Enzymology*, **340**, 450-466.

36. Swalley, S.E., Baird, E.E. and Dervan, P.B. (1997) A pyrrole-imidazole polyamide

motif for recognition of eleven base pair sequences in the minor groove of DNA. *Chem-Eur J*, **3**, 1600-1607.

37. Turner, J.M., Swalley, S.E., Baird, E.E. and Dervan, P.B. (1998) Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J Am Chem Soc*, **120**, 6219-6226.

38. Wang, C.C.C., Ellervik, U. and Dervan, P.B. (2001) Expanding the recognition of the minor groove of DNA by incorporation of beta-alanine in hairpin polyamides. *Bioorganic & Medicinal Chemistry*, **9**, 653-657.

39. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory, Plainview, NY.

40. Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods in Enzymology*, **65**, 499-560.

41. Iverson, B.L. and Dervan, P.B. (1987) Adenine specific DNA chemical sequencing reaction. *Nucleic Acids Research*, **15**, 7823-7830.

42. Warren, C.L., Kartochvil, N.C.S., Hauschild, K.E., Foister, S., Brezinski, M.L., Dervan, P.B., Phillips, G.N., Jr., Ansari, A.Z. (2006) Defining the sequence-recognition profile of DNA-binding molecules. *Proc Natl Acad Sci USA*, **103**, 867-872.