Chapter 2

Conformational Control of DNA Binding Oligomers

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

(Urbach, A. R.; Marques, M. A.; Doss, R. M. and Dervan P. B. "*Controling Oligomer Binding Modes in the DNA Minor Groove"* **2005** – In Preparation)

Abstract.

Polyamides containing imidazole (Im) and pyrrole (Py) amino acids, linked by beta-alanine (β) and gamma-aminobutyric acid (γ), target predetermined DNA sequences in hairpin or extended conformations based on rules for recognition inherent to each binding mode (Figure 2.1).[1,2] In order to address this ambiguity of sequence targeting depending on conformation, we incorporated β, γ, and the α-(R)-amino and α-(R)-acetamido derivatives of γ into the central (X) position of the polyamide Im-β-ImPy-X-Im-β-ImPy-β-Dp (Dp = dimethylaminopropylamine), and examined these compounds in complex with their target hairpin and extended DNA sequences, 5'-TAGTACT-3' and 5'-AAAGAGAAGAG-3', respectively, by DNase I and MPE footprinting. These results demonstrate that hairpin and extended binding modes, which are dependent on ligand conformation, can be controlled by using the appropriate residue at the central position. These design principles should greatly improve the overall sequence fidelity of the next generation of polyamides for DNA recognition in larger, genomic contexts.

Introduction.

Polyamides composed of pyrrole (**Py**), imidazole (**Im**), hydroxypyrrole (**Hp**), and beta alanine (β) amino acids bind to predetermined sequences in the DNA minor groove with affinities and specificities that rival native transcription factors, offering a potentially powerful weapon against human disease.[2] There exist 1:1 and 2:1 ligand-DNA stoichiometries with quite different rules for recognition and, hence, different sequence specificities.[2-5] Pairing rules for recognition in the DNA minor groove have been established that enable discrimination between the four Watson-Crick base pairs using pairs of artificial amino acid residues to specify DNA base pairs.[6-8] Incorporation of the γ-aminobutyric acid (γ) residue promotes folding of the polyamide into a hairpin conformation and substantially increases the affinity and specificity of DNA binding.[9, 10]

Polyamides composed of multiple contiguous heterocyclic residues are overcurved with respect to the DNA helix,[11] and the flexible β residue can be incorporated to better align ligand and DNA, thereby restoring binding affinity.[12, 13] The β residue also promotes extended 1:1 mode of binding for purine-tract DNA using polyamides of type -Im-β-ImPy-β-.[14] Interestingly, the 2:1 and 1:1 binding modes have different rules for DNA recognition.[15] Accordingly, a polyamide may target different DNA sequences, depending on its mode of binding.[16] In an effort to optimize sequence fidelity in larger, genomic contexts, it will be necessary to control the mode of binding (Figure 2.1).

Control over polyamide binding modes has been demonstrated by increasing the length of the central residue by one carbon (β to γ) in order to favor hairpin formation over binding in a 2:1 complex.[17] Solution-phase structural studies showed that both β-

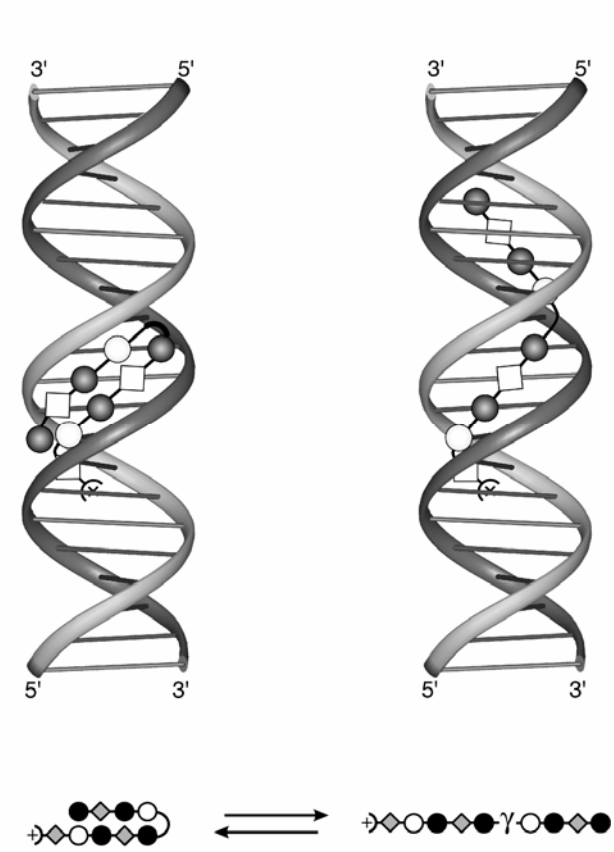

Figure 2.1. Illustration of the equilibrium between hairpin (left) and extended (right) conformational binding modes. The cartoons at top represent DNA with polyamide bound in the respective conformation. DNA is shown as a ladder. Polyamides are shown as dot models, with shaded and non-shaded circles representing imidazole and pyrrole, respectively, and gray diamonds indicating beta-alanine. The gamma turn residue is shown both as a semicircle connecting the two subunits and as the symbol, γ.

and γ-linked polyamides can form hairpins, but that the β-linked compounds are mores sterically strained.[18] Boger and coworkers reported that the hairpin conformation in a β-turn polyamide can be reinforced by substituting the prochiral α-(R)-proton of β with OCH_3.[19] Prior to this work, α-(R)-amino-substituted γ was to increase the binding affinity of hairpin polyamides.[20]

Until now, it has been believed that polyamides containing γ residues should have a strong preference for binding in hairpin conformation. For example, the γ-linked analogue, Im-β-ImPy-**γ**-Im-β-ImPy-β-Dp (**2**), binds with high affinity as a hairpin to its target sequence, 5'-TAGCGCT-3'.[12] However, a similar compound, Im-β-ImPy-**β**-Im-β-ImPy-β-Dp (**1**) (Dp = dimethylaminopropylamide), which differs from **2** by a single methylene unit, binds 5'-AAAGAGAAGAG-3' as a 1:1 complex and 5'-TAGCGCAGCGCTA-3' as a 2:1 complex.[16] Given the structural

similarity of **1** and **2**, it was a simple prediction that **2** may also bind in an extended mode. We find that, indeed, **2** binds to 5'-TAGCGCA-3' in a folded hairpin conformation and to 5'-AAGAGAAGAG-3' in an extended conformation, with surprisingly similar affinities.

This finding prompted us to develop design principles for polyamides that would favor either hairpin or extended binding mode. We hypothesized that the $^{H_2N}\gamma$ residue, which is known to improve hairpin binding,[20] should provide a steric blockade to extended binding. Accordingly, the polyamide Im-β-ImPy-$^{H_2N}\gamma$-Im-β-ImPy-β-Dp (**3**) and its acetamide analogue, Im-β-ImPy-$^{Ac}\gamma$-Im-β-ImPy-β-Dp (**4**), were prepared (Figure 2.2). Based on

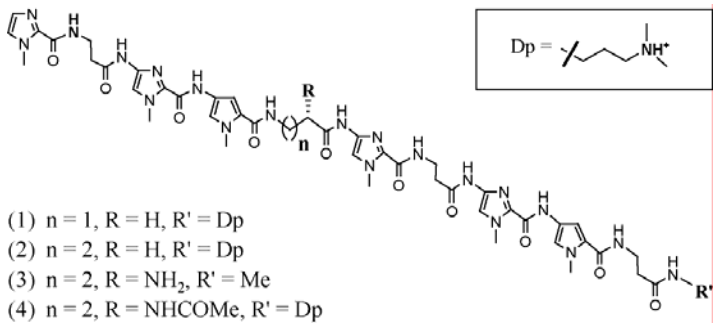

Figure 2.2. Chemical structures of polyamides **1** – **4**.

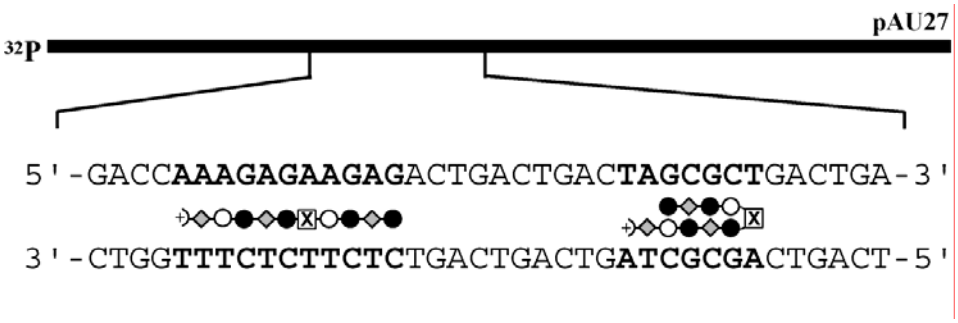

Figure 2.3. The designed insert cloned into plasmid pAU27. The targeted recognition sites are shown in bold type. Polyamides are shown as dot models, with shaded and non-shaded circles representing imidazole and pyrrole, respectively, and gray diamonds indicating beta-alanine. The variable linker position is shown as a square containing the letter X

established recognition rules for the hairpin and extended 1:1 motifs,[1,10,13] we constructed a plasmid containing the target DNA sequences 5'-TAGCGCT-3' (hairpin) and 5'-

AAAGAGAAGAG-3' (extended 1:1) (Figure 2.3). Affinities and binding site sizes for the four compounds in complex with their target sequences were determined in order to compare the relative preferences of these compounds for hairpin and extended binding modes.

Results.

DNA Binding Affinity and Sequence Specificity. Quantitative DNase I footprint titrations[1] were performed for polyamides **1**–**4** on the 288 bp PCR product of pAU27, in order to compare the equilibrium association constants for the resulting complexes. The β-linked compound, **1**, bound to the extended site with very high affinity (K_a = 1.5 x 10^{10} M^{-1}) and more than 150-fold preference over the hairpin site (K_a = 9.7 x 10^7 M^{-1}) (Table 2.1). The γ-linked compound, **2**, bound with similar affinities to the

Table 2.1. Equilibrium Association Constants, K_a (M^{-1})*

Polyamide	Linker	Hairpin 5'-TAGCGCT-3'	Extended 1:1 5'-AAAGAGAAGAG-3'	Specificity (Hairpin/Extended)
1	β	9.7 (±0.8) x 10^7	1.5 (±0.5) x 10^{10}	0.0065
2	γ	7.6 (±0.6) x 10^8	1.3 (±0.5) x 10^8	5.8
3	^{NH_2}γ	1.2 (±0.2) x 10^{10}	7.6 (±0.2) x 10^7	160
4	^{Ac}γ	8.5 (±0.5) x 10^9	1.6 (±0.4) x 10^7	530

* Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. 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.

hairpin (K_a = 7.6 x 10^8 M^{-1}) and extended (K_a = 1.3 x 10^8 M^{-1}) sites. The ^{H_2N}γ-linked polyamide **3** bound to the hairpin site with very high affinity (K_a = 1.2 x 10^{10} M^{-1}) and more than 150-fold specificity versus the extended site (K_a = 7.6 x 10^7 M^{-1}), which is an almost exact reversal of specificity in comparison to compound **1**. Polyamide **4** (^{Ac}γ-linker) also bound with very high affinity to the hairpin site (K_a = 8.5 x 10^9 M^{-1}), but with greater than 500-fold specificity versus the extended site (K_a = 1.6 x 10^7 M^{-1}).

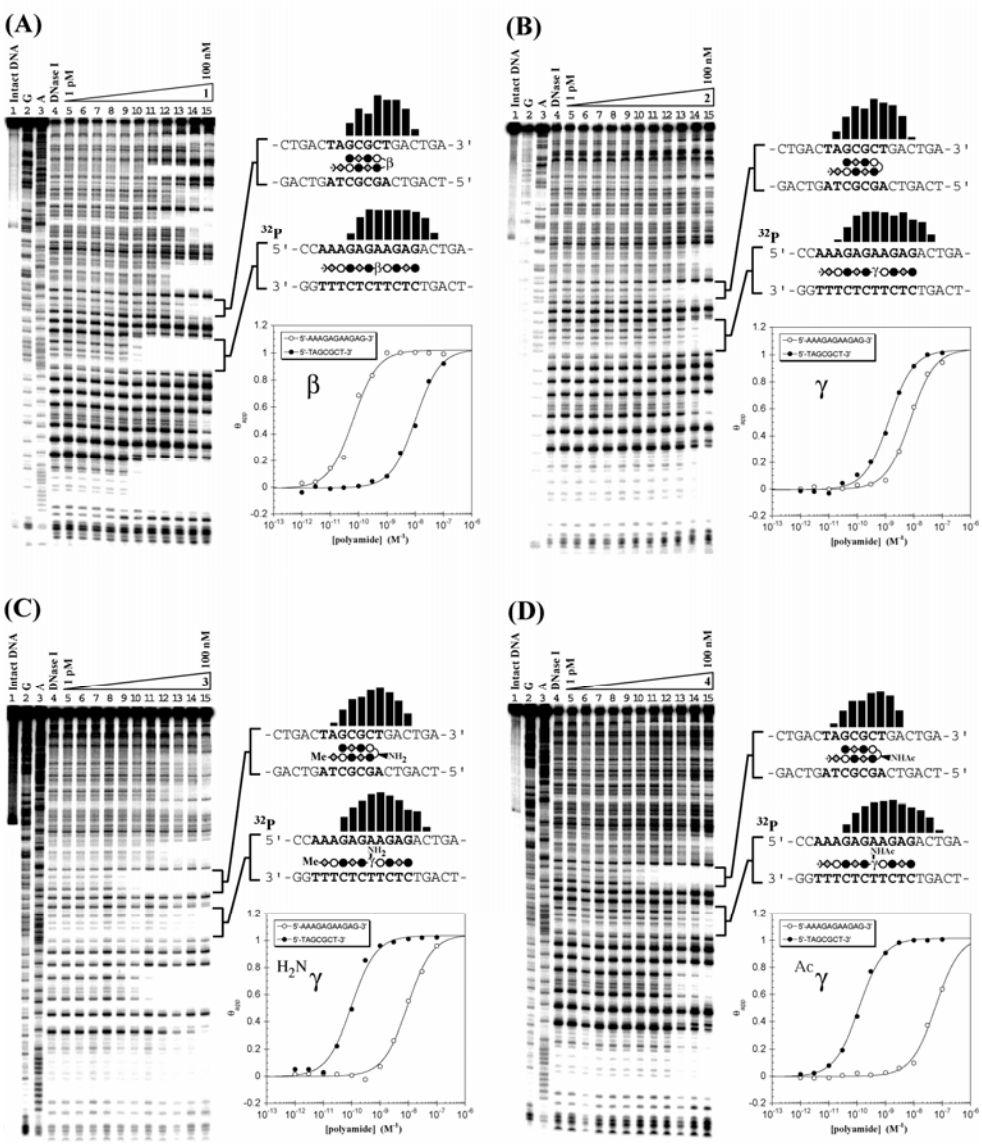

Figure 2.4. (A–D) Quantitative DNase I footprinting[1] for polyamides **1**–**4**, respectively, on the 288 bp, 5'-end-labeled PCR product of plasmid pAU27: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (right top) Schematic illustrating the observed protection pattern derived from the MPE footprinting experiment, with the polyamides shown in the observed conformation; (right bottom) binding isotherms derived from the DNase I footprinting experiment for the two designed sites, 5'-AAAGAGAAGAG-3' and 5'-TAGCGCT-3', as determined from a non-linear least squares fit.

Binding Site Size Determination. Binding modes were deduced from methidium propyl-EDA (MPE) footprinting[1] on the basis of previously characterized hairpin and 1:1 complexes for polyamides **1** and **2** at identical sites.[12, 15, 17] Figure 2.4 displays occupational histograms derived from the MPE gels (not shown), illustrated to the right of the respective DNase gel. In all cases, polyamides bound their target hairpin sites as hairpins and their target extended sites in an extended fashion.

Discussion.

Previous studies of polyamide **1** demonstrated its high affinity for the extended 1:1 binding mode.[15, 16] The >150-fold specificity of **1** for extended versus hairpin binding sites, presented here, may be attributed to the steric destabilization involved in the β residue adopting a hairpin conformation.[18] Moreover, structural studies in 1:1 and 2:1 modes support the notion that β, in an extended conformation, allows for optimal alignment between amino acids residues and DNA base pairs.[13, 21] The γ-linked polyamide, **2**, discriminates least between the hairpin and extended sites, which poses a problem because hairpin polyamides most commonly contain γ as the turn residue. Therefore, a new turn residue that favors hairpin formation over alternative binding modes would be of great value.

Based on molecular modeling using available solution structures,[18, 21] we postulated that incorporation of substituted γ residues should favor hairpin binding while disfavoring extended binding due to steric clashing of the α-(R)-amino group with the wall of the minor groove. Indeed, this result was observed for the ^{H_2N}γ-linked polyamide **3**, which shows >150-fold preference for hairpin versus extended binding—a complete

reversal of preference from compound **1**. It is worth noting that **3** contains a truncated tail in order to maintain a single positive charge. We have also tested the doubly charged Dp-tail analogue of **3** and find that its recognition properties are virtually identical to **3** (data not shown). Compounds **1**–**3** are relatively non-specific, binding with high affinity to many other sites on the plasmid. Compound **4**, which contains the $^{Ac}\gamma$ residue and a Dp tail, binds with high affinity as a hairpin but with markedly reduced non-specific binding. Moreover, **4** exhibits >500-fold preference for hairpin versus extended binding modes, which is an effective 82,000-fold or ~7 kcal/mol reversal of specificity in comparison to **1**. The exceptional specificity of **4** may be attributed to limited mobility of the acetamide group in the hairpin conformation, which should limit alternative binding modes.

Implications for the Design of Minor Groove Binding Polyamides. The results presented here indicate that hairpin and extended modes of polyamide-DNA binding, which are dependent on the ligand conformation, can be controlled by the choice of

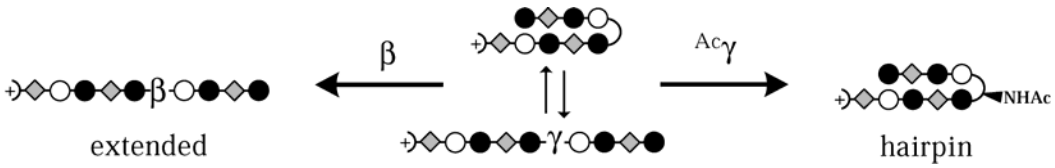

Figure 2.5. Schematic illustrating the relative lack of conformational preference for the γ linker (center) as well as the linkers used to confer a preference for extended (β) and hairpin ($^{Ac}\gamma$) conformations.

linkage between subunits (Figure 2.5). We demonstrate that the compact hairpin binding mode can be favored by incorporating an amino or acetamide substituent at the α-(R) position of γ. The $^{Ac}\gamma$ residue substantially improves the fidelity of hairpin binding, while maintaining high affinity. Alternatively, the extended 1:1 binding mode can be

favored by replacing γ with a β residue. We are currently most interested in favoring hairpin binding because of its higher information density and therefore higher capacity for programmable DNA sequence selection.[3] The design principles elucidated herein should greatly improve the fidelity of sequence recognition for hairpin polyamides in larger, genomic contexts.

Experimental.

General. Methylamine, piperidine, and dimethylaminopropylamine were purchased from Aldrich. Dimethylformamide (DMF) and diisopropylethylamine (DIEA) were purchased from Applied Biosystems. Acetic anhydride and acetonitrile were from EM. Trifluoroacetic acid (TFA) was from Halocarbon. (R)-2-Fmoc-4-Boc-diaminobutyric acid was purchased from Bachem. Boc-β-Pam resin was purchased from Peptides International. HPLC analysis was performed on a Beckman Gold system using a Rainin C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μm C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology and used without further purification. Plasmids were sequenced by Davis Sequencing (Davis, CA). Glycogen (20 mg/mL), dNTP's (PCR nucleotide mix), and all enzymes (unless otherwise stated) were purchased from Boehringer-Mannheim and used with their supplied buffers. pUC19 was from New England Biolabs. Deoxyadenosine [γ-^{32}P] triphosphate was from

ICN. Calf thymus DNA (sonicated, deproteinized) and DNase I (7500 u/mL, FPLC pure) were from Amersham-Pharmacia. AmpliTaq DNA polymerase was from Perkin Elmer and used with the supplied buffers. HEPES was from Sigma. Tris-HCl, dithiothreitol (DTT), RNase-free water (used for all DNA manipulations), and 0.5 M EDTA were from US Biochemicals. Ethanol (200 proof) was from Equistar. Calcium chloride, potassium chloride, and magnesium chloride were from Fluka. Formamide and pre-mixed tris-borate-EDTA (Gel-Mate, used for gel running buffer) were from Gibco. Bromophenol blue was from Acros. All reagents were used without further purification.

Polyamide Synthesis. Polyamides **1**-**4** (Figure 2.2) were prepared according to solid-phase protocols.[22] The synthesis and characterization of polyamides **1** and **2** have been reported previously.[12, 15]

Im-β-ImPy-(R)^{H_2N}γ-Im-β-ImPy-β-Me (**3**). Im-β-ImPy-(R)^{H2N}γ-Im-β-ImPy-β-Pam resin was synthesized in a stepwise fashion from Boc-β-Pam resin (0.59 mmol/g) (Peptides International) using manual solid-phase protocols.[20, 22] The chiral diaminobutyric acid "turn" residue was incorporated by coupling (R)-2-Fmoc-4-Boc-diaminobutyric acid (10 equivalents) to 300 mg Boc-Im-β-Im-Py-β-Pam resin in 2 mL DMF with 1.1 equivalents of DIEA at 37 °C for 2 h, followed by an acetylation wash.[22] Subsequent coupling steps used 1.1 equivalents of DIEA and 45 minute coupling times at room temperature to minimize Fmoc deprotection. Im-β-ImPy-(R)^{Fmoc}γ-Im-β-ImPy-β-Pam resin was treated with piperidine for 20 minutes at room temperature to remove the Fmoc group. 100 mg (38 μmol) of vacuum-dried Im-β-ImPy-(R)^{H_2N}γ-Im-β-ImPy-β-Pam resin was cleaved in 30 mL condensed methylamine in a Parr bomb apparatus at 50 °C

for 2 h, then overnight at room temperature. The methylamine was allowed to evaporate at ambient pressure and temperature, and the resin was suspended in 2 mL acetonitrile, followed by 7 mL 0.1% (wt/v) $TFA_{(aq)}$. The suspension was filtered, and the filtrate was purified by reversed phase preparatory HPLC to afford **3** as a white powder (7.1 μmol, 19% recovery) upon lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1066.49 (1066.48 calcd. for $C_{46}H_{60}N_{21}O_{10}^{+}$).

Im-β-ImPy-(R)^{Ac}γ-Im-β-ImPy-β-Dp (**4**). Im-β-ImPy-(R)^{H_2N}γ-Im-β-ImPy-β-Pam resin was synthesized as described above for **3**. The resin was washed with acetic anhydride in DMF and DIEA for 10 minutes at room temperature and then dried *in vacuo*. 100 mg (37 μmol) Im-β-ImPy-(R)^{Ac}γ-Im-β-ImPy-β-Pam resin was treated with dimethylaminopropylamine at 100 C for 2 h. The resin was removed by filtration, and the filtrate was diluted to 10 mL with 0.1 % (wt/v) $TFA_{(aq)}$ and purified by reversed phase HPLC. **4** was obtained as a white powder (4.0 μmol, 11% recovery) upon lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1179.57 (1179.57 calcd. for $C_{52}H_{71}N_{22}O_{11}^{+}$).

Construction of Plasmid DNA. Plasmid pAU27 (Figure 3) was constructed by inserting the hybridized oligonucleotide, 5'-G A T C C G G G G C C A A A A A G A A A A G A C T G A C T G A C T A G T A C T G A C G A C T G A C C **A A A G A G A A G A G A G** A C T G A C T G A C **T A G C G C T** G A C T G A-3' • 5'-A G C T T C A G T C A G C G C T A G T C A G T C A G T C T C T T C T C T T T G G T C A G T C G T C A G T A C T A G T C A G T C A G T C T T T T C T T T T T G G C C C C G-3', into the *Bam*HI/*Hin*dIII polycloning site in pUC19, with subsequent transformation into

JM109 subcompetent cells (Promega), according to standard protocols.[23] Plasmid DNA was isolated using WizardPlus Midi Preps from Promega.

DNA Radiolabeling and Footprinting Experiments. The 5' end-labeling of plasmid pAU27 as well as DNase I and MPE footprinting experiments were performed exactly in accordance with published protocols.[1] The PCR method was chosen for 5'-end-labeling, employing two primer oligonucleotides, **^{32}P**-5'- A A T T C G A G C T C G G T A C C C G G -3' (forward) and 5'- C T G G C A C G A C A G G T T T C C C G -3' (reverse) to complement the pUC19 *Eco*RI and *Pvu*II sites, respectively, such that amplification by PCR generates the 288-bp, 3'-filled E*co*RI/*Pvu*II restriction fragment.

Acknowledgments.

We are grateful to the National Institutes of Health for financial support, to the Parsons Foundation for fellowship support to M. A. M., and to Caltech for a James Irvine Graduate Fellowship to R. M. D.

References:

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

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

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

[4] 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.

[5] Pelton, J. G.; Wemmer, D. E., *Proceedings of the National Academy of Sciences of the United States of America* **1989,** 86, (15), 5723-5727.

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

[10] Mrksich, M.; Parks, M. E.; Dervan, P. B., *Journal of the American Chemical Society* **1994,** 116, (18), 7983-7988.

[11] 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.

[12] Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1998,** 120, (25), 6219-6226.

[13] de Clairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society* **1999,** 121, (13), 2956-2964.

[14] Janssen, S.; Durussel, T.; Laemmli, U. K., *Molecular Cell* **2000,** 6, (5), 999-1011.

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

[16] Urbach, A. R.; Dervan, P. B., *From Molecular Structure Toward Biology*. ed.; Verlag Helvetica Chimica Acta: Zurich, 2000; p 327-339.

[17] Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B., *Journal of the American Chemical Society* **1996,** 118, (26), 6160-6166.

[18] de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society* **1997,** 119, (34), 7909-7916.

[19] Woods, C. R.; Ishii, T.; Wu, B.; Bair, K. W.; Boger, D. L., *Journal of the American Chemical Society* **2002,** 124, (10), 2148-2152.

[20] Herman, D. M.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1998,** 120, (7), 1382-1391.

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

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

[23] Sambrook, J.; Fritsch, E. F.; Maniatis, T., *Molecular Cloning*. ed.; Cold Springs Harbor Laboratory: Cold Springs Harbor, NY, 1989.