Pergamon

Bioorganic & Medicinal Chemistry 8 (2000) 1947–1955

BIOORGANIC & MEDICINAL CHEMISTRY

Recognition of the DNA Minor Groove by Pyrrole-Imidazole Polyamides: Comparison of Desmethyl- and *N*-Methylpyrrole

Ryan E. Bremer, Jason W. Szewczyk, Eldon E. Baird and Peter B. Dervan*

Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, CA 91125, USA

Received 14 February 2000; accepted 18 April 2000

Abstract—Polyamides consisting of *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methyl-3-hydroxypyrrole (Hp) are synthetic ligands that recognize predetermined DNA sequences with affinities and specificities comparable to many DNA-binding proteins. As derivatives of the natural products distamycin and netropsin, Py/Im/Hp polyamides have retained the *N*-methyl substituent, although structural studies of polyamide:DNA complexes have not revealed an obvious function for the *N*-methyl. In order to assess the role of the *N*-methyl moiety in polyamide:DNA recognition, a new monomer, desmethylpyrrole (Ds), where the *N*-methyl moiety has been replaced with hydrogen, was incorporated into an eight-ring hairpin polyamide by solid-phase synthesis. MPE footprinting, affinity cleavage, and quantitative DNase I footprinting revealed that replacement of each Py residue with Ds resulted in identical binding site size and orientation and similar binding affinity for the six-base-pair (bp) target DNA sequence. Remarkably, the Ds-containing polyamide exhibited an 8-fold loss in specificity for the match site versus a mismatched DNA site, relative to the all-Py parent. Polyamides with Ds exhibit increased water solubility, which may alter the cell membrane permeability properties of the polyamide. The addition of Ds to the repertoire of available monomers may prove useful as polyamides are applied to gene regulation in vivo. However, the benefits of Ds incorporation must be balanced with a potential loss in specificity.

Introduction

Cell-permeable small molecules with the ability to target predetermined DNA sequences and interfere with gene expression would be valuable tools in molecular biology and potentially in human medicine. Recent efforts in our laboratories have been inspired by the natural products netropsin and distamycin, crescent-shaped di- and tripeptides composed of *N*-methylpyrrole (Py) carboxamides that recognize four or five successive A•T base pairs.[1–8] Distamycin and netropsin normally form 1:1 complexes in the minor groove of DNA.[9–12] However, structural studies have revealed that at high concentrations distamycin forms an antiparallel side-by-side dimer tightly packed into the minor groove, with each monomer making hydrogen bonds from the backbone amides to the bases on one DNA strand.[13,14] Atomic substitutions to the *N*-methylpyrrole scaffold of distamycin have provided new aromatic amino acids, *N*-methylimidazole (Im) and *N*-methyl-3-hydroxypyrrole (Hp), that allow for the recognition of predetermined DNA sequences.[15–17] A set of pairing rules have been developed that correlate each Watson–Crick base pair with a side-by-side pairing of the Py, Hp, and Im aromatic amino acids (Fig. 1).[16–18] An antiparallel pairing of Im opposite Py (Im/Py pair) distinguishes G•C from C•G and both of these from A•T and T•A base pairs.[16,18–21] A Py/Py pair binds both A•T and T•A in preference to G•C and C•G.[13,14,16,18–20,22,23] The discrimination of T•A from A•T using Hp/Py and Py/Hp pairs, respectively, completes the four base pair (bp) code.[17,24,25] The linker amino acid γ-aminobutyric acid (γ) connects the polyamide subunits in a 'hairpin motif' and enhances affinity >100-fold relative to the unlinked dimers.[26–29] A C-terminal β-alanine (β) increases both affinity and specificity and facilitates solid-phase synthesis.[27,30] The γ and β residues both favor recognition of A•T and T•A base pairs.[31] Each polyamide subunit recognizes the minor groove with a N–C orientation relative to the 5′-3′ direction of the DNA helix.[21,32] Eight-ring hairpin polyamides with affinities and specificities comparable to those of DNA-binding proteins have been shown to permeate a variety of cell types in culture and specifically interfere with gene expression.[33–35]

Extensive biochemical studies have revealed the incremental contributions to DNA affinity and specificity of many features of polyamides, most importantly the effects of substitutions at the 3-position of the aromatic

*Corresponding author. Tel.: +1-626-356-6002; fax: +1-626-683-8753; e-mail: dervan@caltech.edu

0968-0896/00/$ - see front matter
PII: S0968-0896(00)00145-0

rings.[16–18,22] However, in every case, the *N*-methyl substituent present in the distamycin and netropsin has been retained. Structural studies do not reveal an obvious role for the *N*-methyls in polyamide:DNA recognition, as they make no DNA contacts and point away from the minor groove, toward the solvent.[20,24,29,36] As natural products, distamycin and netropsin may have faced selective pressures based on many factors, including biosynthesis, solubility, permeability, and DNA affinity and specificity. It remained to be determined if the *N*-methyl is a substituent that can be discarded, or if it plays an important role in polyamide:DNA recognition, in particular in the side-by-side pairings. This question provided the impetus to examine a desmethyl-(*N*-methylpyrrole) ring, that replaces the *N*-methyl group of the pyrrole with hydrogen. Desmethyl-(*N*-methylpyrrole) should follow the polyamide pairing rules similarly to *N*-methylpyrrole. The sequence specific elements of Py/Im/Hp polyamides are the two atoms at the C-3 position of each ring pair, whilst the N-1 position projects out of the minor groove and does not make sequence specific DNA contacts. For the sake of a two-letter code, we have routinely referred to *N*-methylimidazole and *N*-methylpyrrole residues as imidazole and pyrrole with the abbreviations Im and Py, respectively. Thus, for clarity, the formal desmethyl-(*N*-methylpyrrole) residue will be referred to as desmethylpyrrole and it will be represented in the two-letter code as Ds.

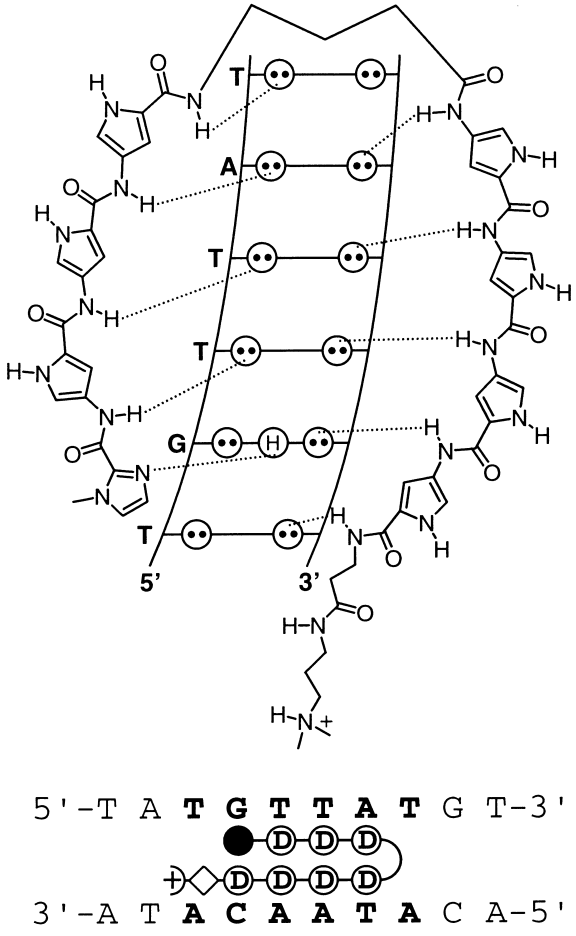
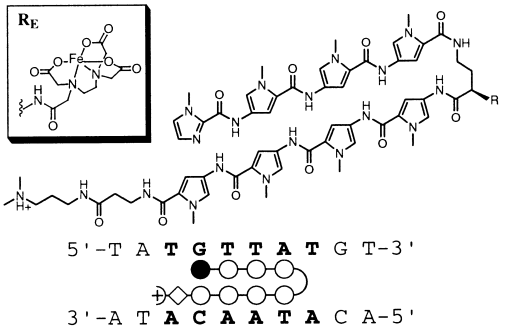


Figure 1. (Top) Hydrogen bonding model of ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**) bound in the minor groove to the 5′-TGTTAT-3′ match site. Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. (Bottom) Model of **2** bound to the 5′-TGTTAT-3′ match site (bold). *N*-Methylimidazole and desmethylpyrrole are represented by black circles and white circles with a 'D' inside, respectively. The γ-turn is shown as a curved line. β-Alanine and Dp are depicted as a diamond and a plus sign, respectively.

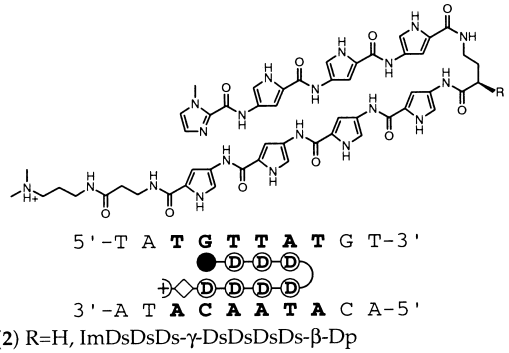
Distamycin analogues with single desmethylpyrrole substitutions exhibited enhanced antiviral and cytotoxic properties relative to distamycin.[37,38] It was postulated that the enhanced activity derived from increased cell permeability for the desmethylpyrrole analogues.[37] The potential benefits of desmethylpyrrole substitutions on biological activity provided further motivation to examine the DNA binding energetics of Ds-containing hairpin polyamides.

Here we report the DNA binding properties of two eight-ring hairpin polyamides: a control polyamide containing all Im/Py and Py/Py pairs, ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**), and a polyamide completely substituted with Ds, ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**) (Fig. 2). Binding site size and location were determined by MPE footprinting, while affinity cleavage revealed binding site orientation and stoichiometry.[4,8,39] Quantitative DNase I footprint titrations were employed to measure the



(**1**) R=H, ImPyPyPy-γ-PyPyPyPy-β-Dp

(**1-E**) R=R_E, ImPyPyPy-(*R*)$^{EDTA\cdot Fe(II)}$γ-PyPyPyPy-β-Dp



(**2**) R=H, ImDsDsDs-γ-DsDsDsDs-β-Dp

(**2-E**) R=R_E, ImDsDsDs-(*R*)$^{EDTA\cdot Fe(II)}$γ-DsDsDsDs-β-Dp

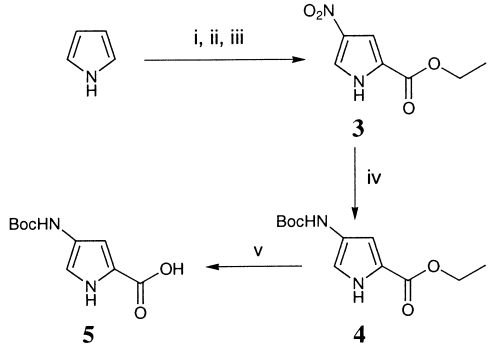
Figure 2. Structures of eight-ring hairpin polyamides ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**), and ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**), and the corresponding EDTA-modified derivatives, **1-E** and **2-E**. Schematics of each polyamide bound to the 5′-TGTTAT-3′ match site are depicted as in Figure 1, with an open circle representing *N*-methylpyrrole.

equilibrium association constants (K_a) of each polyamide at its match and mismatch sites.[40–42] Finally, not only will the Ds monomer reveal the role of the *N*-methyl substituent in polyamide:DNA recognition, but each incorporation will decrease the polyamide molecular weight and add a hydrogen bond donor that can interact with the solvent, potentially increasing polyamide solubility in water and altering cell permeability, as well.

Results

Desmethylpyrrole monomer and polyamide synthesis

Analogues of distamycin have been prepared previously by solution phase methods with a single desmethylpyrrole substitution using the Boc-Ds-acid (**5**).[37] It remained to be determined if a hairpin polyamide fully substituted with desmethylpyrrole in place of each pyrrole residue could be synthesized by solid-phase protocols.[30] Towards this end, we report here a convenient synthesis of **5** based on the preparation of the *N*-methylpyrrole derivative that provides **5** in 25 g scale without column chromatography (Scheme 1).[30] Polyamide **1** has been previously reported.[33] The solid-phase synthesis of desmethylpyrrole containing polyamides, **2**, was performed on β-Ala-Pam resin using DCC/HOBt activated **5**, Boc-Py-OBt, and Boc-Im-acid, according to previously reported procedures.[30] Cleavage of the resin bound polyamide by aminolysis with dimethylaminopropylamine (Dp) afforded polyamide **2**, following reversed phase HPLC purification. Although the recovery of polyamide **2** was lower than that generally observed for Py-containing polyamides, the solid-phase protocols did provide Ds-containing polyamides of equal purity to the Py-containing polyamides. For the synthesis of affinity cleavage analogues with EDTA, the desired polyamides were synthesized using (*R*)-2-F_{moc}-4-Boc-diaminobutyric acid in place of Boc-γ-aminobutyric acid.[32] After aminolysis and reversed phase HPLC purification, the deprotected 2-amino group was treated with an excess of EDTA-dianhydride, and the remaining anhydride hydrolyzed to provide **1-E** and **2-E**, following another reversed phase HPLC purification.[32] Polyamides were characterized by analytical reversed phase HPLC, ^{1}H NMR, and MALDI-TOF mass spectroscopy. Polyamide **2** showed a significantly reduced retention time in reversed phase HPLC (21.9 min), relative to the all-Py analogue, **1** (28.8 min).



Scheme 1. Synthesis of the Boc-Ds-acid monomer (**5**) for solid-phase synthesis: (i) trichloroacetyl chloride, ethyl ether; (ii) HNO_3 (fuming), H_2SO_4, acetic anhyride; (iii) NaH, ethanol; (iv) (a) H_2 (1 atm), 10% Pd/C, DMF; (b) Boc-anhyride, DIEA; (v) 1 M KOH, MeOH, water, 60 °C.

Identification of binding sites by MPE·Fe(II) footprinting

MPE·Fe(II) footprinting was performed on the 3′- and 5α-^{32}P end-labeled 279 bp *Eco*RI/*Pvu*II restriction fragment of pSES4 (Figs 3 and 4).[4] On the basis of the polyamide pairing rules, the sites 5′-TGTTAT-3′ and 5′-TGGTAT-3′ are for **1** and **2** 'match' and 'single-base-pair mismatch' sites, respectively (mismatch underlined). The footprint cleavage protection patterns of **1** and **2** revealed the polyamides bound the expected match site, 5′-TGTTAT-3′ (Fig. 5A and B). The location and size of the 3′-shifted footprints is consistent with the eight-ring polyamides binding as hairpins in the minor groove to the expected 6 bp target. A MPE cleavage protection pattern was also observed at the mismatch site, with what appeared to be a potentially elongated footprint to the 3′-side of the bottom strand. Subsequent affinity cleavage experiments revealed the significance of this observation (vide infra). A number of additional footprints on the restriction fragment corresponding to match sequences of the form 5′-(A,T)G(A,T)$_4$-3′ and a single-base-pair mismatch have been identified, but their MPE footprints were not quantitated, due to the closeness of the bands in this region of the gel.

Identification of binding orientation and stoichiometry by affinity cleavage

Affinity cleavage polyamides, **1-E** and **2-E** were used to determine binding orientation and stoichiometry for the Ds polyamides.[8,39] The EDTA·Fe(II) moiety of polyamides **1-E** and **2-E** was attached via the γ-turn. The lack of a flexible linker results in a narrow cleavage pattern centered on the position of the γ-turn. Affinity cleavage titrations were performed on the 3′- and 5′-^{32}P end-labeled *Eco*RI/*Pvu*II restriction fragment of pSES4 (Fig. 4). A single cleavage locus at the 3′-side of the 5′-TGTTAT-3′ site indicated each polyamide bound the match site as a monomer in a single orientation (Fig. 5C and D). The expected cleavage pattern at the 3′-side of the 5′-TGGTAT-3′ single base pair mismatch site was also observed for **1-E** and **2-E**. An unexpected cleavage locus between the designed match and mismatch sites was believed to result from the polyamide bound in a reverse orientation (3′–5′, N–C) with the γ-turn centered on the

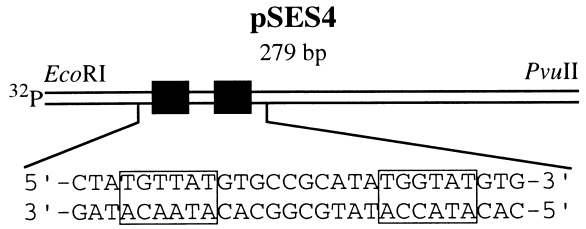


Figure 3. The 279 bp *Eco*RI/*Pvu*II restriction fragment of pSES4. The match and single-base-pair mismatch sequences, 5′-TGTTAT-3′ and 5′-TGGTAT-3′ (mismatch underlined), respectively, are boxed.

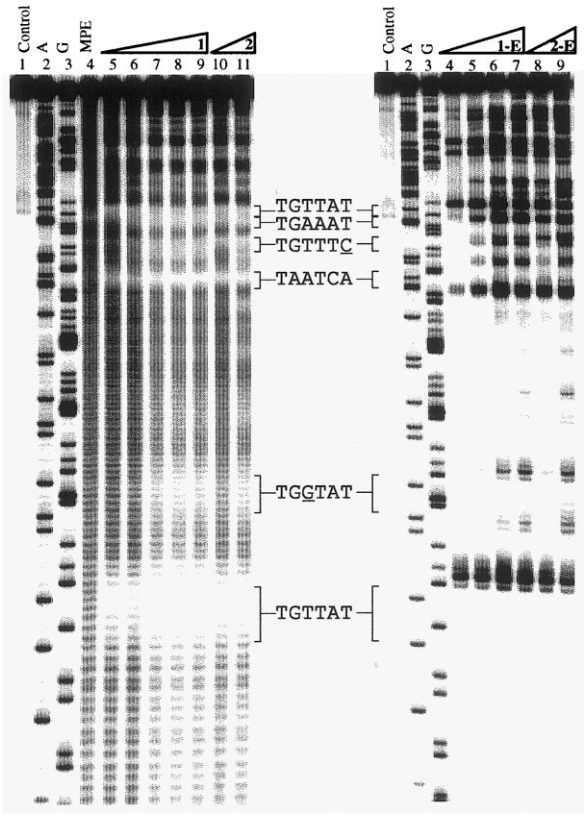


Figure 4. Storage phosphor autoradiograms of MPE•Fe(II) footprinting (left)[4] and affinity cleavage experiments (right)[8,39] on the 5′-^{32}P end-labeled 279 bp *Eco*RI/*Pvu*II restriction fragment of pSES4. Match and single-base-pair mismatch sites (mismatch underlined) are indicated in the middle. MPE•Fe(II) footprinting (left): lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, MPE•Fe(II) standard; lanes 5–9, 10, 100, 200, 500 nM, 1 μM, **1**; lanes 10–11, 10, 100 nM, **2**. All reactions were performed in a final volume of 400 μL containing 20 kcpm 5′-radiolabeled DNA, 20 mM HEPES (pH 7.3), 200 mM NaCl, 50 μg/mL glycogen, 0.5 μM MPE•Fe(II), and 5 mM DTT. Linear storage phosphor range 160–1600. Affinity cleavage (right): lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4–7, 1, 10, 100 nM, 1 μM **1-E**; lanes 8–9, 1, 10 nM **2-E**. All reactions were performed in a final volume of 400 μL containing 20 kcpm 5′-radiolabeled DNA, 20 mM HEPES (pH 7.3), 200 mM NaCl, 50 μg/mL glycogen, 0.5 μM Fe(II)$(NH_4)_2SO_4$, and 5 mM DTT. Linear storage phosphor range 200–2000.

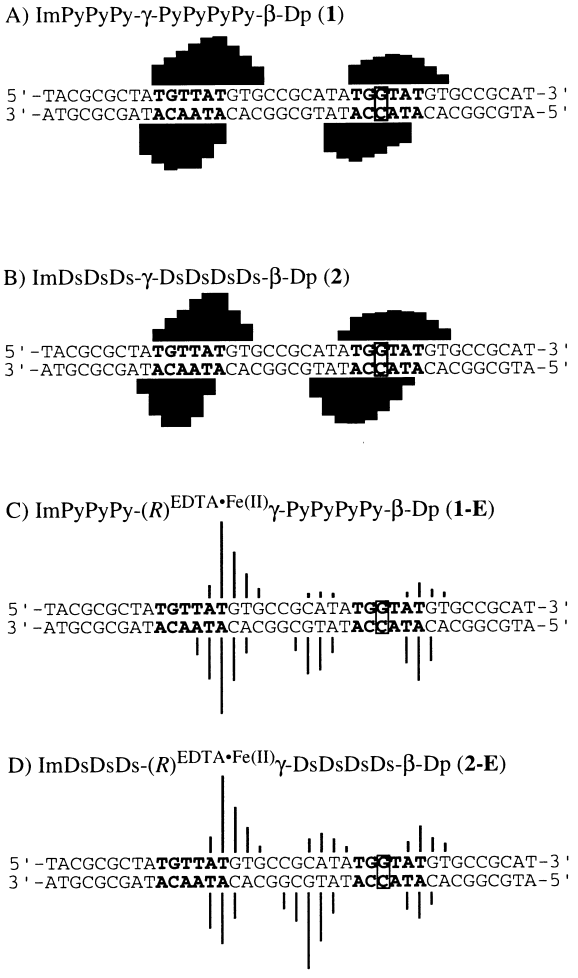


Figure 5. (A-B) MPE•Fe(II) protection patterns with (A) ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) (500 nM), and (B) ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**) (100 nM). Bar heights are proportional to the realvtive protection from cleavage for each band. (C) and (D) Affinity cleavage patterns of (C) ImPyPyPy-γ-PyPyPyPy-β-Dp-EDTA•Fe (**1-E**) 1 μM, and (D) ImDsDsDs-γ-DsDsDsDs-β-Dp-EDTA•Fe (**2**) (10 nM). Line heights are proportional to the cleavage intensity of each band. The 6-bp polyamide binding site 5′-TGTTAT-3′ and the designed single-base-pair mismatch 5′-TGGTAT-3′are shown in bold.

bold **T** in 5′-ATATGG-3′. This binding mode is also a single base pair mismatch according to the pairing rules, because it places the β-alanine residue over a C•G base pair. Binding to the designed mismatch and the reverse orientation mismatch was observed for both the Py- and Ds-containing polyamides.

Quantitative DNase I footprint titrations

Quantitative DNase I footprint titrations were performed on the 3′-^{32}P end-labeled *Eco*RI/*Pvu*II restriction fragment of pSES4 (Fig. 6).[40–42] Equilibrium association constants (K_a) for each polyamide bound to the match and mismatch sites were determined by calculating a fractional saturation value at both sites for each polyamide concentration and fitting the data to a modified form of the Hill equation (Fig. 7). Both the Py- and Ds-containing eight-ring hairpin polyamides bound the 5′-TGTTAT-3′ target site with similar affinity: **1**, $1.2\pm0.4\times10^{10}\,M^{-1}$; **2**, $1.5\pm0.2\times10^{10}\,M^{-1}$ (Table 1). The presence of overlapping binding sites for the designed mismatch and the reverse orientation mismatch, precludes assignment of the resulting footprint to a specific binding mode. However, the relative affinity of **1** and **2** at this combined mismatch footprint is revealing. Polyamide **1** demonstrated >200-fold preference for the match site versus the mismatch ($5.9\pm0.9\times10^{7}\,M^{-1}$). Ds-containing polyamide **2** demonstrated a 25-fold preference for its match site versus the mismatch ($5.9\pm2.3\times10^{8}\,M^{-1}$); an 8-fold loss in specificity relative to **1**.

In addition to the designed six-base-pair match (5′-TGTTAT-3′) and mismatch (5′-TGGTAT-3′) sites, there are other binding sites seen in the MPE footprinting experiment (Fig. 4), which are distal to the ^{32}P label: a match site, 5′-TGATTA-3′, two contiguous match sites,

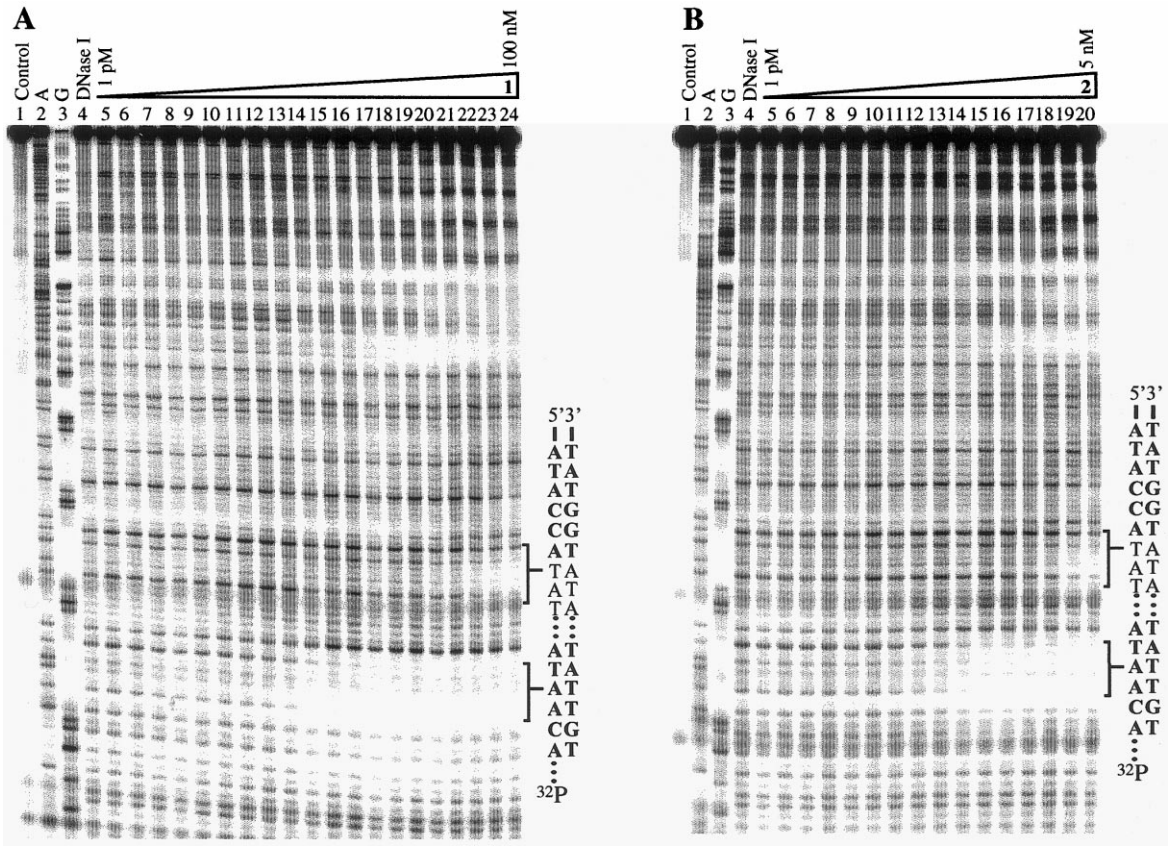


Figure 6. Storage phosphor autoradiograms of quantitative DNase I footprint titration experiments[40–42] with (A) ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) and (B) ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**), on the 3′-^{32}P-end-labeled 279 bp *Eco*RI/*Pvu*II restriction fragment from pSES4. All reactions contained 20 kcpm restriction fragment, 10 mM Tris–HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ and were performed at 22 °C. Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5–20, 1, 2, 5, 10, 15, 25, 40, 65, 100, 150, 250, 400, 650 pM, 1, 2, 5 nM, respectively of **1** or **2**. Footprinting gels of **1** contained the additional lanes 21–24 of 10, 20, 50, 100 nM, respectively. the positions of the match and mismatch sequences are indicated. The designed match (5′-TGTTAT) and single base pair mismatch (5′-TGGTAT-3′) mismatch sites are in bold. The overlapping reverse orientation mismatch site observed by affinity cleavage (5′ATATGG-3′) is also labeled. Linear storage phosphor ranges for A and B were 30–500 and 80–1000, respectively.

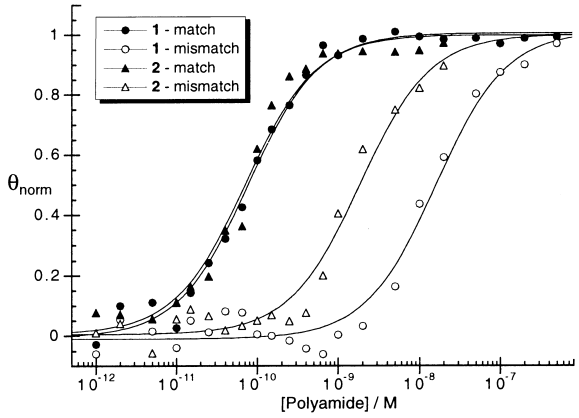


Figure 7. Data from the quantitative DNase I footprint titration experiments for ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**), and ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**) bound to the match and mismatch sites. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data for the binding of **1** to the match and mismatch is indicated by filled and unfilled circles, respectively. The data for the binding of **2** to the match and mismatch is indicated by filled and unfilled triangles, respectively. The solid curves are the best-fit binding titration isotherms obtained from nonlinear least squares algorithm.[27]

5′-TGAAATGTTAT-3′, and a single-base-pair mismatch site, 5′-TGTTTC-3′. We were able to quantitate binding of **1** and **2** to 5′-TGATTA-3′ and 5′-TGAAATGTTAT-3′ by DNase I footprinting at the concentration ranges used here and for the sake of completeness, their equilibrium association constants are reported (Table 2).

Discussion

In the reported syntheses of desmethylpyrrole analogues of distamycin, the N-1 position proved reactive under the coupling conditions, resulting in side reactions and lowering the yield of the desired product.[37] The solid-phase synthesis of Ds-containing polyamides also suffered from lower recoveries than the Py-containing analogues, suggesting that future application of Ds may require protection of the N-1 position.

MPE•Fe(II) footprinting and affinity cleavage studies revealed that replacement of the *N*-methyl of the pyrrole carboxamide with hydrogen had little effect on polyamide binding site size, orientation or stoichiometry at the 6-bp target site. The footprint protection patterns

Table 1. Equilibrium association constants (M^{-1})[a]

Polyamide	Match[b]	Mismatch[c]	Specificity[d]
ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**)	1.2×10^{10}	5.9×10^{7}	203
ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**)	1.5×10^{10}	5.9×10^{8}	25

[a]Values reported are the mean values obtained from three DNase I footprint titration experiments. The assays were carried out at 22 °C, 10 mM Tris–HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
[b]The 'match' affinity is for the polyamide binding the 5′-TGTTAT-3′ target site.
[c]The 'mismatch' affinity is derived from the footprint arising from the binding of the overlapping reverse orientation and single base pair mismatch sites (see text).
[d]Specificity is defined as K_a(match)/K_a(mismatch).

Table 2. Equilibrium association constants (M^{-1})[a]

Polyamide	5′-caTGATT-Acg-3′	5′-tgTGAAAT-GTTATcc-3′
ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**)	$1.7\ (\pm0.2)\times10^{9}$	$2.2\ (\pm0.1)\times10^{9}$
ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**)	$3.3\ (\pm1.1)\times10^{9}$	$6.7\ (\pm1.6)\times10^{9}$

[a]Values reported are the mean values obtained from three DNase I footprint titration experiments. The assays were carried out at 22 °C, 10 mM Tris–HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

and cleavage loci at the match site were consistent with 1:1 polyamide:DNA complexes bound as hairpins in the minor groove. The expected affinity cleavage pattern was also observed at the 3′-side of the designed single base pair mismatch site. Affinity cleavage revealed the polyamide exhibited an additional unexpected binding mode assigned as reverse orientation (3′–5′, N–C) binding to 5′-ATATGG-3′. The MPE footprints of **1** and **2** at the mismatch site are consistent with polyamide binding primarily at the designed mismatch site, 5′-TGGTAT-3′, with only a slightly expanded footprint on the bottom strand perhaps resulting from polyamide binding to the reverse orientation site. Polyamides have demonstrated >10-fold preference for binding with the polyamide oriented N–C along the 5′–3′ direction of the targeted DNA strand.[21] Furthermore, this orientation preference would be expected to be enhanced by the *R*-enantiomer of the affinity cleavage analogues (**1-E** and **2-E**), which would direct the bulky EDTA moiety into the floor of the minor groove if the polyamide bound in the reverse orientation.[32] Thus, it was surprising that **1-E** and **2-E** bound both the designed mismatch and the reverse orientation mismatch with similar relative affinities, as revealed by affinity cleavage. It is important to note that the incorporation of Ds resulted in enhanced binding relative to **1** for both of these mismatch modes, as evidenced by cleavage at *both* mismatch sites at lower concentration for **2** (10 nM) than **1** (1 μM) (Fig. 4). In other words, the removal of the *N*-methyl substituent does not result in a loss of specificity versus one mismatch binding mode, but rather versus both mismatches observed here.

As expected, polyamide **1** bound the 5′-TGTTAT-3′ site with subnanomolar affinity and excellent specificity. Remarkably, replacing all of the *N*-methylpyrroles with desmethylpyrrole in **2** did not have a significant effect on the polyamide's DNA binding affinity to the target site. Polyamide **2** bound the target site with similar affinity to **1**, revealing the *N*-methyl group is not necessary for subnanomolar binding by hairpin polyamides. However, removal of the *N*-methyl substituents resulted in an 8-fold loss in specificity for **2**, relative to **1**. The overlapping binding modes at the mismatch site precludes the development of a model based on the structure of the polyamide:DNA complex to account for the higher affinity at the mismatch site exhibited by Ds-containing polyamides. However, in the DNase I footprint titrations, the mismatch footprint clearly appears at approximately 10-fold lower polyamide concentrations for **2** than **1**. The affinity cleavage results strongly suggest that this DNase I footprint arises from a combination of both mismatch binding modes.

It was unexpected that incorporation of Ds would result in differential polyamide–DNA interactions at the match and mismatch sites. Further studies, including structural studies of Ds-containing polyamides bound to match and mismatch sites, will be necessary to elucidate the basis for the decrease in specificity upon Ds-introduction. The *N*-methyl moiety in the ring–ring pairs may be necessary to keep the pyrrole rings set deeply in the minor groove. In the absence of the *N*-methyl, the energetic penalty may be decreased for the exocyclic amine of guanine pushing the ring away from the DNA; allowing for an increased tolerance of G•C base pairs. The *N*-methyl may also play a role in aligning the side-by-side register of the ring pairings. Furthermore, it may be possible for some or all of the Ds carboxamides to rotate about the amide backbone and place the N-1 toward the floor of the minor groove. The implications on specificity of such a structural alteration at the polyamide:DNA interface may be significant.

Conclusions

The introduction of desmethylpyrrole into the hairpin polyamide motif has provided insight into the role of the *N*-methyl substituent in polyamide:DNA recognition. While the *N*-methyl is not necessary for subnanomolar binding by hairpin polyamides, it is clearly important for attaining optimal target site specificity. It is now becoming clear that a complete understanding of polyamide:DNA recognition will require not only structural studies of a variety of polyamides bound to their DNA targets, but of equal importance will be an analysis of the structural basis of specificity achieved by analyzing structures of polyamides bound to mismatch DNA sites. Current efforts to develop larger polyamides that bind longer sequences may face difficulty with issues of solubility and cell permeability. Desmethylpyrrole containing polyamides exhibit higher solubility than their *N*-methylpyrrole analogues and may potentially alter polyamide cell permeability, making them useful alternatives in the design of second-generation polyamides to recognize larger binding sites for genomic applications. In the future, the benefits of each incorporation of Ds must be balanced with the potential loss in sequence specificity.

Experimental

All synthetic reagents were as previously described.[30] Analytical HPLC was performed on a Beckman Gold Nouveau system with a model 126 pump and model 168 diode array detector. A Rainen C_{18}, Microsorb MV, 5 μm, 300×4.6 mm reverse phase column was employed with 0.1% (w/v) TFA:H_2O and 1.5% acetonitrile/min. Preparatory HPLC was performed on a Hamilton PRP-1 reversed phase C_{18}, 250×21.5 mm, 12–20 μm column with 0.5% (w/v) acetic acid and 0.56%/min acetonitrile unless otherwise noted. Resin substitution of synthesized polyamides was calculated as $L_{new}(mmol/g) = L_{old}/(1 + L_{old}(W_{new} - W_{old}) \times 10^{-3})$ where L is the loading (mmol of amine per g of resin) and W is the weight (g mol^{-1}) of the growing peptide attached to the resin.[30] DNA restriction fragment labeling, DNase I footprinting, and determination of equilibrium association constants were accomplished using previously described protocols.[41,42] Chemical sequencing reactions were performed according to published methods.[43,44] Plasmid pSES4 was previously prepared by S. Swalley using reported procedures with the following inserts: 5′-GATCGGCTACGCGCTATGTTATGTGCCGCATATGGTATGTGCCGCATATGGGATGTGCGCCATATGGGGGTGTGCCG-3′ and 5′-AGCTCGGCACACCCCATATGGCGCACATCCCATATGCGGCACATACCATATGCGGCACATAACATAGCGCGTAGCC.[31,42]

Monomer synthesis

Ethyl 4-nitropyrrole-2-carboxylate (3). 2-(trichloroacetyl) pyrrole was available in kg quantities by scaling the literature procedure.[45] Nitration of 2-(trichloroacetyl) pyrrole was accomplished using the previously reported procedure for 1-methyl-2-(trichloroacetyl)pyrrole with minor alterations.[30] The addition of nitric acid was performed at −20 °C and the reaction was quenched with water. Esterification of the 2-(trichloroacetyl)-4-nitropyrrole was accomplished using the procedure reported for the *N*-methyl analogue without reflux.[30] The solid isolated after precipitation with water was recrystalized from ethyl acetate:hexanes:methanol to provide a light tan powder (551 g, 3.0 mol, 22% yield from pyrrole): ^{1}H NMR (DMSO-d_6) δ 13.13 (s, 1H), 8.01 (d, 1H, J = 1.5 Hz), 7.20 (d, 1H, J = 1.9 Hz), 4.25 (q, 2H, J = 7.2 Hz), 1.27 (t, 3H, J = 7.2 Hz); EIMS m/e 184.0482 (184.0484 calcd for $C_7H_8N_2O_4$).

Ethyl 4-[(*tert*-butoxycarbonyl)amino]pyrrole-2carboxylate (4). Ethyl 4-nitropyrrole-2-carboxylate (128 g, 0.75 mol) was dissolved in DMF (640 mL) and 23 g of 10% Pd/C added. The mixture was stirred under H_2 (50 psi) for 2 h. The Pd/C was removed by filtration through Celite, followed by washing with DMF (100 mL). Di-*tert*-butyl dicarbonate (164 g, 0.75 mol) and triethylamine (315 mL, 2.26 mol) was added and stirred for 1.25 h. The mixture was poured into 1 L of water and extracted with diethyl ether. The ether was passed over solid sodium bicarbonate, decanted, dried ($MgSO_4$) and concentrated to in vacuo to provide a white solid (125 g, 0.49 mol, 65% yield): ^{1}H NMR (DMSO-d_6) δ 11.50 (s, 1H), 9.08 (s, 1H), 6.96 (s, 1H), 6.59 (s, 1H), 4.18 (q, 2H, J = 7.2 Hz), 1.42 (s, 9H), 1.24 (t, 3H, J = 7.1 Hz); FABMS m/e (254.1266 calcd for $C_{12}H_{18}N_2O_4$).

4-[(*tert*-Butoxycarbonyl)amino]pyrrole-2 carboxylic acid (5). Ethyl 4-[(*tert*-butoxycarbonyl)amino]pyrrole-2carboxylate (31.0 g, 122 mmol) was dissolved in ethanol (124 mL); 1 M KOH (620 mL) was added and the solution heated at 65 °C for 4 h. The solution was cooled to room temperature and the ethanol removed in vacuo. The solution was diluted with water (800 mL) and extracted with diethyl ether. The pH of the aqueous layer was reduced to ca. 3 with 10% (v/v) H_2SO_4 and the mixture extracted with diethyl ether. The combined ether extracts were dried ($MgSO_4$) and concentrated in vacuo to provide a white powder (25.6 g, 113 mmol, 93% yield): ^{1}H NMR (DMSO-d_6) δ 12.17 (s, 1H), 11.34 (s, 1H), 9.03 (s, 1H), 6.90 (s, 1H), 6.53 (s, 1H), 1.42 (s, 9H); FABMS m/e 249.0841 (M + Na 249.0851 calcd for $C_{10}H_{14}N_2NaO_4$).

Polyamide synthesis

ImDsDsDs-γ-DsDsDsDs-β-Dp (2). ImDsDsDs-γ-DsDs DsDs-β-PAM-resin was synthesized in a stepwise fashion by manual solid-phase methods from Boc-β-PAM-resin (600 mg, 0.75 mmol/g).[30] A sample of polyamide resin (217 mg, 0.46 mmol/g) was cleaved with dimethylaminopropylamine (2 mL, 55 °C, 18 h). The crude polyamide was diluted to 8 mL with 0.1% (w/v) TFA then purified by reverse phase HPLC using a Waters DeltaPak 25× 100 mm 100 μM C_{18} column in 0.1% (w/v) TFA, gradient elution 0.25%/min. CH_3CN. ImDsDsDs-γ-DsDs-DsDs-β-Dp was recovered upon lyophilization of the appropriate fractions as a white powder (1.7 mg, 1.5 μmol, 1.5% recovery); UV (H_20) λ_{max} 312 (66,000); analytical reverse phase HPLC rt = 21.9 min; ^{1}H NMR (DMSO-d_6): δ 11.28 (s, 1H), 11.20 (s, 4H), 11.11 (s, 2H), 10.42 (s, 1H), 9.93 (s, 2H), 9.88 (s, 3H), 9.82 (s, 1H), 9.2 (br s, 1H), 8.04 (m, 3H), 7.35 (s, 1H), 7.23 (s, 1H), 7.17 (s, 3H), 7.14 (s, 1H), 7.13 (s, 1H), 7.06 (s, 1H), 7.04 (s, 1H), 7.02 (s, 2H), 7.01 (s, 2H), 6.90 (s, 1H), 6.83 (s, 1H), 6.81 (s, 1H), 3.95 (s, 3H), 3.61 (m, 2H), 3.19 (m, 2H), 3.06 (quartet, 2H, J = 5.8 Hz), 2.93 (quintet, 2H, J = 4.6 Hz), 2.68 (d, 6H, J = 4.6 Hz), 2.2 (br m, 4H), 1.92 (m, 2H), 1.7 (br m, 2H); MALDI-TOF-MS (monoisotopic) [M + H] 1123.6, 1123.5 calcd for $C_{52}H_{59}N_{20}O_{10}^+$.

ImDsDsDs-(R)^{EDTA}γ-DsDsDsDs-β-Dp (2-E). ImDsDs Ds-(R)^{NH2}γ-DsDsDsDs-β-PAM-resin was synthesized in a stepwise fashion by manual solid-phase methods from Boc-β-PAM-resin (600 mg, 0.75 mmol/g).[30] A sample of polyamide resin (215 mg, 0.46 mmol/g) was cleaved with dimethylaminopropylamine (2 mL, 55 °C, 18 h). The crude polyamide was diluted to 8 mL with 0.1% (w/v) TFA then purified by reverse phase HPLC using a Waters DeltaPak 25×100 mm 100 μM C_{18} column in 0.1% (w/v) TFA, gradient elution 0.25%/min. CH_3CN. ImDsDsDs-(R)^{NH2}γ-DsDsDsDs-β-Dp was recovered upon lyophilization of the appropriate fractions as a white powder (2.1 mg, 1.9 μmol, 1.9% recovery); UV (H_2O) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 11.42 (s, 1H), 11.20 (m, 5H), 11.11 (s, 1H), 10.60 (s, 1H), 10.41 (s, 1H), 9.95 (s, 1H), 9.92 (s, 3H), 9.87 (s, 1H), 9.2 (br s, 1H), 8.27 (m, 2H), 8.04 (m, 1H), 7.35 (s, 1H), 7.23 (s,

1H), 7.17 (s, 4H), 7.12 (s, 1H), 7.02 (m, 6H), 6.95 (s, 1H), 6.90 (s, 1H), 6.82 (s, 1H), 3.95 (s, 3H), 3.31 (m, 2H), 3.20 (s, 2H), 3.06 (quartet, 2H, $J = 5.6$ Hz), 2.93 (quintet, 2H, $J = 4.9$ Hz), 2.69 (d, 6H, $J = 4.5$ Hz), 2.2 (br m, 4H), 1.95 (m, 2H), 1.64 (m, 2H); MALDI-TOF-MS (monoisotopic) [M + H] 1138.5, 1138.5 calcd for $C_{52}H_{60}N_{21}O_{10}^{+}$. ImDsDsDs-$(R)^{NH2}$γ-DsDsDsDs-β-Dp (700 nmol) was functionalized with EDTA as described[32] (68 nmol, 9.7% recovery) MALDI-TOF-MS (monoisotopic) [M + H] 1412.7, 1412.6 calcd for $C_{62}H_{74}N_{23}O_{17}^{+}$.

ImPyPyPy-$(R)^{EDTA}$γ-PyPyPyPy-β-Dp (1-E). ImPyPy Py-$(R)^{NH2}$γ-PyPyPyPy-β-PAM-resin was synthesized in a stepwise fashion by manual solid-phase methods from Boc-β-PAM-resin (600 mg, 0.75 mmol/g).[30] A sample of polyamide resin (350 mg, 0.44 mmol/g) was cleaved with dimethylaminopropylamine (2 mL, 55 °C, 18 h). The crude polyamide was diluted to 8 mL with 0.1% (w/v) TFA then purified by reverse phase HPLC. ImPyPyPy-$(R)^{NH2}$γ-PyPyPyPy-β-Dp was recovered upon lyophilization of the appropriate fractions as a white powder (25.3 mg, 20.5 μmol, 13.3% recovery); UV (H_2O) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.40 (s, 1H), 9.93 (s, 1H), 9.92 (s, 1H), 9.89 (s, 2H), 9.87 (s, 2H), 9.75 (br s, 1H), 8.05 (m, 1H), 7.97 (m, 1H), 7.84 (t, 1H, $J = 5.4$ Hz), 7.35 (s, 1H), 7.24 (d, 1H, $J = 0.9$ Hz), 7.20 (s, 2H), 7.16 (s, 1H), 7.14 (s, 2H), 7.12 (m, 2H), 7.01 (s, 2H), 6.99 (s, 2H), 6.92 (s, 1H), 6.83 (s, 1H), 6.78 (s, 1H), 3.94 (s, 3H), 3.80 (s, 12H), 3.75 (m, 9H), 3.23 (m, 5H), 3.00 (m, 2H), 2.23 (m, 2H), 2.16 (m, 2H), 2.04 (m, 2H), 2.03 (s, 6H), 1.42 (m, 2H); MALDI-TOF-MS (monoisotopic) [M + H] 1236.7, 1236.6 calc'd for $C_{59}H_{74}N_{21}O_{10}^{+}$. ImPyPyPy-$(R)^{NH2}$γ-PyPyPyPy-β-Dp (6.4 mg, 5.2 μmol) was functionalized with EDTA as described[32] (3.5 mg, 2.3 μmol, 43% recovery) MALDI-TOF-MS (monoisotopic) [M + H] 1510.9, 1510.7 calcd for $C_{69}H_{88}N_{23}O_{17}^{+}$.

MPE footprinting and affinity cleavage

All reactions were carried out in a final volume of 400 μL. A polyamide stock solution (or water for reference lanes) was added to an assay buffer where the final solution conditions were as follows: 20 mM HEPES (pH 7.3), 200 mM NaCl, 50 μg/mL glycogen and 20 kcpm 3′- or 5′-end-labeled *Eco*RI/*Pvu*II fragment of pSES4. The reactions were equilibrated for 16 h at 22 °C. For MPE footprinting, MPE·Fe(II) was added to a final concentration of 0.5 μM and equilibrated for 5 min.[4] For affinity cleavage experiments, $Fe(NH_4)_2(SO_4)_2$ was added to a final concentration of 0.5 μM and equilibrated for 20 min.[8,39] Cleavage was initiated by the addition of DTT to a final concentration of 5 mM and allowed to proceed for 30 min. Reactions were terminated with ethanol (1 mL) and 10 μL of a precipitation buffer (2.8 mg/mL glycogen, 140 μM bp calf thymus DNA). The labeled DNA was precipitated and analyzed on a 8% denaturing polyacrylamide gel as described for DNase I footprinting.[41,42]

Acknowledgements

We are grateful to the National Institutes of Health for research support, the National Science Foundation, Bristol-Myers Squibb and the Ralph M. Parsons Foundation for predoctoral fellowships to R.E.B., the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B., and the National Institutes of Health for a research service award to J.W.S.

References and Notes

1. Arcamone, F.; Penco, S.; Prezzi, P. G.; Nicolella, V.; Pirelli, A. *Nature* **1964**, *203*, 1064.
2. Zasedatelev, A. S.; Gursky, G. V.; Zimmer, C.; Thrum, H. *Mol. Biol. Rep* **1974**, *1*, 337.
3. Krylov, A. S.; Grokhovsky, S. L.; Zasedatelev, A. S.; Zhuze, A. L.; Gursky, G. V.; Gottikh, B. P. *Nucleic Acids Res.* **1979**, *6*, 289.
4. Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470.
5. Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res.* **1983**, *11*, 5555.
6. Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 9271.
7. Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861.
8. Dervan, P. B. *Science* **1986**, *232*, 464.
9. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553.
10. Coll, M.; Frederick, C. A.; Wang, A. H. J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385.
11. Klevit, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986**, *25*, 3296.
12. Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1988**, *27*, 8088.
13. Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723.
14. Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Nat. Struct. Biol.* **1994**, *1*, 169.
15. Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688.
16. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.
17. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.
18. White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569.
19. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586.
20. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104.
21. White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756.
22. Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J. *Biochemistry* **1999**, *38*, 2143.
23. Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393.
24. Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111.
25. Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 11621.
26. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
27. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147.
28. Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306.
29. deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909.
30. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.

31. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc*. **1999**, *121*, 1113.
32. Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc*. **1998**, *120*, 1382.
33. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
34. Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.
35. Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890.
36. deClairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc*. **1999**, *121*, 2956.
37. Grehn, L.; Ragnarsson, U.; Eriksson, B.; Oberg, B. *J. Med. Chem*. **1983**, *26*, 1042.
38. Grehn, L.; Ragnarsson, U.; Datema, R. *Acta Chem. Scand. B* **1986**, *40*, 145.
39. Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457.
40. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol*. **1986**, *130*, 132.
41. Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc*. **1997**, *119*, 7636.
42. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. — Eur. J*. **1997**, *3*, 1600.
43. Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res*. **1987**, *15*, 7823.
44. Maxam, A. M.; Gilbert, W. S. *Methods Enzymol*. **1980**, *65*, 499.
45. Bailey, D. M.; Johnson, R. E.; Albertson, N. F. *Org. Synth* **1988**, *50*, 618.