Toward rules for 1:1 polyamide:DNA recognition

Adam R. Urbach and Peter B. Dervan*

Division of Chemistry and Chemical Engineering, 164-30, California Institute of Technology, Pasadena, CA 91125

Contributed by Peter B. Dervan, February 12, 2001

Polyamides composed of four amino acids, imidazole (Im), pyrrole (Py), hydroxypyrrole (Hp), and β-alanine (β), are synthetic ligands that form highly stable complexes in the minor groove of DNA. Although specific pairing rules within the 2:1 motif can be used to distinguish the four Watson·Crick base pairs, a comparable recognition code for 1:1 polyamide:DNA complexes had not been described. To set a quantitative baseline for the field, the sequence specificities of Im, Py, Hp, and β for the four Watson·Crick base pairs were determined for two polyamides, Im-β-ImPy-β-Im-β-ImPy-β-Dp (1, for Im, Py, and β) and Im-β-ImHp-β-Im-β-ImPy-β-Dp (2, for Hp), in a 1:1 complex within the DNA sequence context 5′-AAAGAGAAGAG-3′. Im residues do not distinguish G,C from A,T but bind all four base pairs with high affinity. Py and β residues exhibit ≥10-fold preference for A,T over G,C base pairs. The Hp residue displays a unique preference for a single A·T base pair with an energetic penalty.

At the forefront of the endeavor to control gene expression by small molecules is the elucidation of chemical principles for direct read-out of predetermined sequences of double-stranded DNA. Polyamide ligands containing the aromatic amino acids pyrrole (Py), imidazole (Im), and hydroxypyrrole (Hp) form 1:1 and 2:1 complexes with DNA (1–3). The 2:1 motif has proven versatile, whereby pairs of amino acids can be assembled side-by-side to distinguish all four Watson·Crick base pairs (Im/Py = G·C; Py/Im = C·G; Hp/Py = T·A; and Py/Hp = A·T) (3). Py-Im polyamides are overcurved with respect to the DNA helix (4–6). By substituting the aliphatic amino acid β-alanine (β) for Py, the ligand curvature is reset, allowing specific recognition of 16 contiguous base pairs of DNA (7–9). Dimeric complexes are capable of binding in slipped modes as well as fully paired overlapped modes (8). Covalent coupling of the dimer creates hairpin and cycle ligands that lock the antiparallel subunits into place but at the cost of doubling the ligand size (10, 11). With the goal of targeting larger binding site sizes, one is led to consider the potential limitation of ligand size on cell-permeability.

Although there exists a significant body of literature on the 1:1 mode of binding, particularly by Lown and coworkers (12), much of this research was carried out before quantitative footprinting methods were introduced to the field. In an effort to characterize more rigorously the 1:1 mode of binding, we address the question quantitatively whether polyamides in a 1:1 complex can discriminate any of the four Watson·Crick base pairs. Our goal is to establish an energetic baseline for future studies. For example, Laemmli and coworkers reported recently that β-linked polyamides of sequence type Im-β-ImPy-β-Im-β-Im-β-Dp bind tracts of 5′-GAGAA-3′ repeats with 1:1 stoichiometry in a single orientation (13). For the study here, a similar polyamide Im-β-ImPy-β-Im-β-ImPy-β-Dp, **1**, is chosen as the template to examine the specificity at Im, Py, and β residues in an oriented 1:1 complex with DNA. Because Hp/Py pairs were shown to discriminate T·A from A·T in the 2:1 motif, a second ligand Im-β-ImHp-β-Im-β-ImPy-β-Dp, **2,** was prepared to explore any possible specificity the Hp residue may have in a 1:1 complex with DNA.

Specificity at a single and unique carboxamide position was determined by varying a single base pair within the parent sequence context, 5′-AAAGAGAAGAG-3′, to all four Watson·Crick base pairs to compare the relative affinities for the four possible complexes. To meet this end, three plasmids were cloned, each containing four binding sites: pAU8 (for Im), 5′-AAAGA**X**AAGAG-3′; pAU15 (for β), 5′-AAAG**X**GAAGAG-3′; and pAU12 (for Py and Hp) 5′-AAAGAGA**X**GAG-3′, where **X** = A, T, G, and C (Fig. 1). To examine the effects of multiple base pair mutations on polyamide binding, the plasmid pAU18 was prepared, which contains the four different binding sites, 5′-AAA**X**A**X**AAA**X**A**X**AAA-3′ (**X** = G, C, A, T). Quantitative DNase I footprint titrations were used to characterize the energetics of each complex, and affinity cleavage experiments revealed the orientation preference at each site.

Materials and Methods

General. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology and used without further purification. Plasmids were sequenced by the Sequence/Structure Analysis Facility (SAF) at the California Institute of Technology. Glycogen (20 mg/ml), dNTPs (PCR nucleotide mix), and all enzymes (unless otherwise stated) were purchased from Boehringer-Mannheim 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 units/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, DTT, RNase-free water (used for all DNA manipulations), and 0.5 M EDTA were from United States Biochemical. Ethanol (200 proof) was from Equistar. Calcium chloride, potassium chloride, and magnesium chloride were from Fluka. Formamide and premixed 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. DNA manipulations were performed according to standard protocols (14).

Polyamides. Polyamides **1** and **2** were synthesized by solid phase methods (15, 16) (Fig. 2). Polyamide **1** has been characterized (17).

Im-β-ImHp-β-Im-β-ImPy-β-Dp (2). The hydroxypyrrole-containing polyamide, **2**, is prepared by using the methyl ether-protected hydroxypyrrole amino acid, Boc-Op-OH. The protected polyamide, Im-β-ImOp-β-Im-β-ImPy-β-Dp, was synthesized in a stepwise fashion by manual solid-phase methods from Boc-β-Pam-Resin (0.24 mmol/g) (15). A sample of Im-β-ImOp-β-Im-β-ImPy-β-Pam-Resin (300 mg, 0.19 mmol/gram) was placed in a glass 20-ml peptide synthesis vessel and treated with neat dimethylaminopropylamine (2 ml) at 55°C with periodic agitation for 16 h. The reaction mixture was then filtered to remove resin, 0.1% (wt/vol) trifluoroacetic acid was added (6 ml), and the resulting solution was purified by reversed phase HPLC. Im-β-ImOp-β-Im-β-ImPy-β-Dp is recovered on lyophilization of the appropriate fractions as a white powder (9.1 mg, 14%

*To whom reprint requests should be addressed. E-mail: dervan.caltech.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

BIOCHEMISTRY

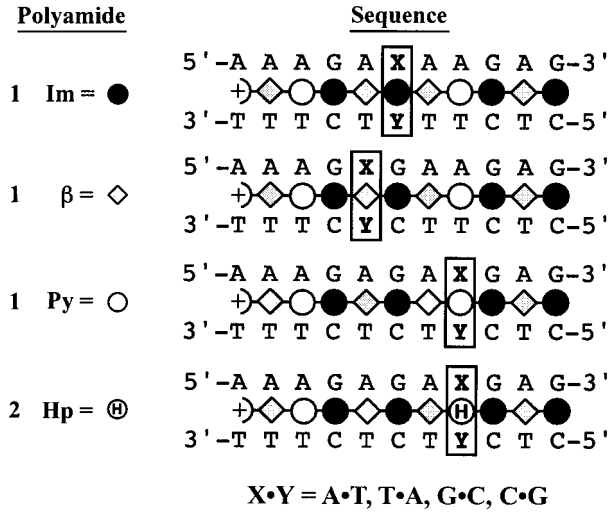

Fig. 1. Schematic illustrating the examination of sequence selectivity at a single imidazole (Im), beta alanine (*β*), pyrrole (Py), or hydroxypyrrole (Hp) position within the parent context, 5′-AAAGAGAAGAG-3′. Imidazole and pyrrole rings are represented as filled and open circles, respectively; *β*-alanines are shown as gray diamonds; and hydroxypyrrole is indicated by a circle containing the letter H.

recovery). UV (H_2O) λmax 260 (ε ≈ 28,000); ^{1}H NMR (DMSO-d_6) δ 10.50 (s, 1H), 10.42 (s, 1 H), 10.33 (s, 1 H), 9.91 (s, 1 H), 9.27 (br s, 1 H), 8.91 (s, 1 H), 8.41 (t, 1 H, J = 6.0 Hz), 8.04 (t, 2 H, J = 6.0 Hz), 7.89 (t, 1 H, J = 6.0 Hz), 7.49 (s, 1 H), 7.42 (s, 1 H), 7.41 (s, 1 H), 7.34 (s, 1 H), 7.19 (s, 1 H), 7.14 (s, 1 H), 6.99 (s, 1 H), 6.92 (s, 1 H), 3.29 (s, 3 H), 3.92 (m, 6 H), 3.90 (s, 3 H), 3.78 (s, 3 H), 3.75 (s, 3 H), 3.69 (s, 3 H), 3.49 (m, 6 H), 3.36 (q, 2 H, J = 5.3 Hz), 3.09 (q, 2 H, J = 6.3 Hz), 2.97 (t, 2 H, J = 5.7 Hz), 2.72 (d, 6 H, J = 4.8 Hz), 2.58 (m, 6 H), 2.33 (t, 2 H, J = 6.9 Hz), 1.72 (quintet, 2 H, J = 8.4 Hz); matrix-assisted laser desorption ionization time of flight (MALDI-TOF)-MS (monoisotopic), 1138.58 (1138.54 calculated for $C_{50}H_{68}N_{21}O_{11}^+$).

To remove the methoxy protecting group, a sample of Im-*β*-ImOp-*β*-Im-*β*-ImPy-*β*-Dp (9 mg, 7.9 μmol) was treated with sodium thiophenoxide at 100°C for 2 h as described previously (16). Dimethylformamide (1.0 ml) and thiophenol (0.5 ml) were placed in a (13 × 100 mm) disposable Pyrex screw cap culture tube. A 60% dispersion of sodium hydride in mineral oil (100 mg) was slowly added. On completion of the sodium hydride addition, Im-*β*-ImOp-*β*-Im-*β*-ImPy-*β*-Dp (5 mg) dissolved in dimethylformamide (0.5 ml) was added. The solution was agitated, and heated to 100°C for 2 h. On completion of the reaction, the reaction mixture was cooled to 0°C, and 7 ml of a 20% (wt/vol) aqueous solution of trifluoroacetic acid added. The aqueous layer is separated from the resulting biphasic solution. To remove the last trace of thiophenol, the aqueous layer was extracted three times with diethyl ether. The deprotected polyamide was purified by reversed phase preparatory HPLC. Im-*β*-ImHp-*β*-Im-*β*-ImPy-*β*-Dp **2** is recovered on lyophilization of the appropriate fractions as a white powder (1.5 mg, 18% recovery). UV (H_2O) λmax 260 (ε ≈ 28,000), ; ^{1}H NMR (DMSO-d_6) δ 10.38 (s, 1 H), 10.37 (s, 1 H), 10.33(s, 1 H), 9.93 (s, 1 H), 9.90 (s, 1 H), 9.61 (s, 1 H), 9.18 (br s, 1 H), 8.37 (t, 1 H, J = 6.0 Hz), 8.04 (m, 2 H), 7.90 (t, 1 H, J = 6.0 Hz), 7.56 (t, 1 H, J = 5.9 Hz), 7.48 (s, 1 H), 7.43 (s, 1 H), 7.40 (s, 1 H), 7.31 (s, 1 H), 7.19 (s, 1 H), 7.12 (s, 1 H), 6.94 (s, 1 H), 6.93 (s, 1 H), 3.92 (m, 9 H), 3.90 (s, 3 H), 3.78 (s, 3 H), 3.74 (s, 3 H), 3.48 (m, 6 H), 3.09 (q, 2 H, J = 5.7 Hz), 2.99 (m, 6 H), 2.56 (m, 4 H), 2.48 (d, 6 H, J = 5.4 Hz), 2.32 (t, 2 H, J = 5.9 Hz), 1.61 (quintet, 2 H, J = 5.4 Hz); MALDI-TOF-MS (monoisotopic), 1124.6 (1124.5 calcd. for $C_{49}H_{66}N_{21}O_{11}^+$).

Construction of Plasmid DNA. Plasmids were constructed by inserting the following hybridized inserts into the *Bam*HI/*Hin*DIII polycloning site in pUC19.

pAU8. 5′-GATCCGGCCAAAGA**G**AAGAGAAACCGGG-GCCAAAGA**C**AAGAGAAACCGGGGCCAAAGA**A**AA-GAGAAACCGGGGCCAAAGA**T**AAGAGAAACCGGA-3′·5′-AGCTTCCGGTTTCTCTT**A**TCTTTGGCCCCGGTTT-CTCTT**T**TCTTTGGCCCCGGTTTCTCTT**G**TCTTTGGCC-CCGGTTTCTCTT**C**TCTTTGGCCG-3′.

pAU12. 5′-GATCCGGCCAAAGAGA**A**GAGAAACCGG-GGCCAAAGAGA**T**GAGAAACCGGGGCCAAAGAGA**G**-GAGAAACCGGGGCCAAAGAGA**C**GAGAAACCGGA-3′·5′-AGCTTCCGGTTTCTC**G**TCTCTTTGGCCCCGGTT-TCTC**C**TCTCTTTGGCCCCGGTTTCTC**A**TCTCTTTGGC-CCCGGTTTCTC**T**TCTCTTTGGCCG-3′.

pAU15. 5′-GATCCGGCCAAAGA**G**AAGAGAAACCGG-GGCCAAAG**T**GAAGAGAAACCGGGGCCAAAG**G**GAA-GAGAAACCGGGGCCAAAG**C**GAAGAGAAACCGGA-3′·5′-AGCTTCCGGTTTCTCTTC**G**CTTTGGCCCCGGTTT-CTCTTC**C**CTTTGGCCCCGGTTTCTCTTC**A**CTTTGGCC-CCGGTTTCTCTTC**T**CTTTGGCCG-3′.

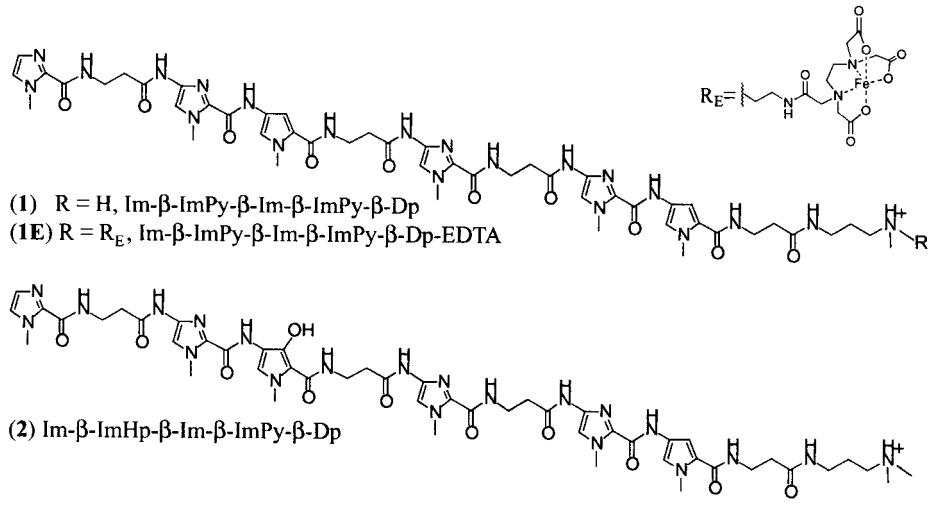

Fig. 2. Structures of polyamides **1**, **1E**, and **2** with their polyamide sequences.

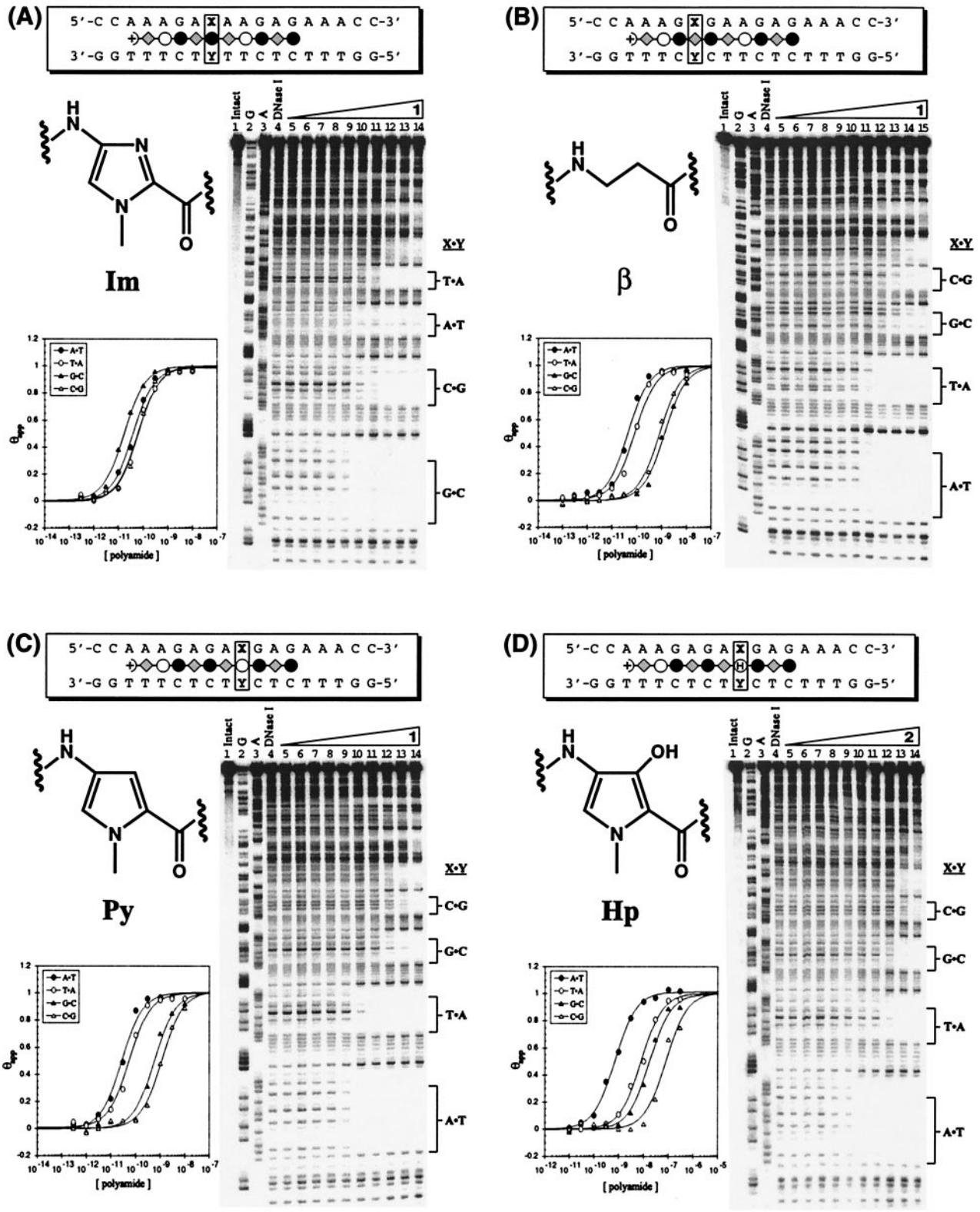

Fig. 3. Quantitative DNase I footprint titration experiments for polyamide 1 on the 298-bp, 5′-end-labeled PCR product of plasmids pAU8 (*A*), pAU15 (*B*), and pAU12 (*C*), as well as polyamide **2** on the same fragment of pAU12 (*D*). (*A* and *C*) Lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5–14, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM **1**, respectively. (*B*) Lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5–15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM **1**, respectively. (*D*) Lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5–14, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM, 300 nM **2**, respectively. Each footprinting gel is accompanied by the following: (*Top*) Binding schematic with the mutated position boxed; (*Middle Left*) chemical structure of the monomer of interest; and (*Bottom Left*) binding isotherms for the four designed sites. θ_{norm} values were obtained according to published methods (8). The data points for the solid lines are best-fit Langmuir binding titration isotherms obtained by a nonlinear, least-squares algorithm.

Table 1. Equilibrium association constants, K_a, M^{-1}

X·Y	Im	β	Py	Hp
A·T	2.5 (±0.2) × 10^{10}	2.4 (±0.1) × 10^{10}	3.4 (±0.3) × 10^{10}	1.6 (±0.2) × 10^{9}
T·A	1.1 (±0.1) × 10^{10}	1.3 (±0.1) × 10^{10}	1.5 (±0.2) × 10^{10}	1.3 (±0.2) × 10^{8}
G·C	2.6 (±0.4) × 10^{10}	4.3 (±1.4) × 10^{8}	1.8 (±0.2) × 10^{9}	4.9 (±0.8) × 10^{7}
C·G	1.3 (±0.3) × 10^{10}	7.8 (±1.9) × 10^{8}	8.6 (±1.5) × 10^{8}	1.0 (±0.3) × 10^{7}

pAU18. 5′-GATCCGGCCAAA**G**A**G**AA**G**A**G**AAACCGG-GGCCAAA**C**A**C**AA**C**A**C**AAACCGGGGCCAAA**A**AA-AA**A**AAAACCGGGGCCAAA**T**AT**A**A**T**AT**A**AACCGGA-3′·5′-AGCTTCCGGTTT**A**TA**T**TA**T**ATTTGGCCCCGGT-TT**T**T**T**T**T**T**T**TTTGGCCCCGGTTT**G**T**G**TT**G**T**G**TTTGG-CCCCGGTTT**C**T**C**TT**C**T**C**TTTGGCCG-3′.

Preparation of 5′-End-Labeled Restriction Fragments. The two primer oligonucleotides, 5′-AATTCGAGCTCGGTAC-CCGG-3′ (forward) and 5′-CTGGCACGACAGGTTTC-CCG-3′ (reverse) were constructed to complement the pUC19 *Eco*RI and *Pvu*II sites, respectively, such that amplification by PCR generates the 298-bp, 3′-filled *Eco*RI/*Pvu*II restriction fragment. The forward primer was radiolabeled by using [γ-^{32}P]dATP and polynucleotide kinase. The PCR product was purified on a 7% non-denaturing preparatory polyacrylamide gel (5% cross-link) and isolated. Chemical sequencing reactions were performed according to published protocols (18, 19).

Affinity Cleavage (20). All reactions were carried out in a volume of 400 μl. A polyamide stock solution or water (for reference lanes) was added to an assay buffer, where the final concentrations were 28.6 mM Hepes, 85.7 mM NaCl buffer (pH 7.0), 200 mM NaCl, and 20,000 cpm 3′- or 5′-radiolabeled DNA. The solutions were allowed to equilibrate for 12 h. A fresh solution of ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$; 1 μM) was added to each equilibrated solution (except intact), and the resulting mixture was allowed to stand for 15 min. Cleavage was initiated by the addition of 10 μl of 5 mM DTT and allowed to proceed for 30 min. Reactions were stopped by precipitation in 1 ml ethanol and 10 μl of a 140 μM bp calf thymus DNA/2.8 mM glycogen solution, and the pellets were washed in 75% ethanol, dried *in vacuo*, resuspended in 15 μl H_2O, lyophilized to dryness, and resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer. The solutions were then denatured at 90°C for 8 min, and a 5-μl sample (≈8 kcpm) was loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) and run at 2000 V.

Quantitative DNase I Footprint Titrations (21). All reactions were carried out in a volume of 400 μl. We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. A polyamide stock solution (or water for reference and intact lanes) was added to an assay buffer where the final concentrations were 10 mM Tris·HCl buffer (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 20,000 cpm 5′-radiolabeled DNA. The solutions were allowed to equilibrate for 18 h at 22°C. Cleavage was initiated by the addition of 10 μl of a DNase I solution (diluted with 1 mM DTT to 1.5 units/ml) and allowed to proceed for 7 min at 22°C. The reactions were stopped by adding 50 μl of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/ml glycogen, and 30 μM base-pair calf thymus DNA, and then ethanol precipitated (2.1 volumes) at 14,000 rpm for 23 min. The pellets were washed with 75% ethanol, resuspended in 15 μl RNase-free water, lyophilized to dryness, and then resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer (with bromophenol blue as dye), denatured at 90°C for 10 min, and loaded directly onto a prerun 8% denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 1.2 h. The gels were dried *in vacuo* at 80°C and then exposed to a storage phosphor screen (Molecular Dynamics). Equilibrium association constants were determined as previously described (8).

Results

Sequence Specificity. Quantitative DNase I footprint titrations were carried out for polyamides Im-β-ImPy-β-Im-β-ImPy-β-Dp (**1**) and Im-β-ImHp-β-Im-β-ImPy-β-Dp (**2**) on three different restriction fragments. Sequence specificity at a single position was determined by varying a single base pair within the parent sequence context, 5′-AAAGAGAAGAG-3′, to all four Watson·Crick base pairs to directly compare the equilibrium association constants (Figs. 2 and 3). The variable base pair position was chosen opposite the amino acid residue in question. For specificity at Im, polyamide **1** bound the four sites within the sequence, 5′-AAAGA**X**AAGAG-3′ (**X** =A, T, G, C), with similar high affinities, K_a = 2.6–1.1 × 10^{10} M^{-1}, revealing that Im accepts all four base pairs (Table 1). For specificity at the Py or β positions, polyamide **1** bound the sites, 5′-AAAGAGA**X**-GAG-3′ and 5′-AAAG**X**GAAGAG-3′, respectively (**X** = A, T, G, C), with high affinity and in both cases displays a preference for A·T or T·A > G·C or C·G by a factor of 10. Remarkably, substituting one Py residue with Hp afforded the most specific polyamide that bound the sequence, 5′-AAAGAGA**X**GAG-3′, with A·T > T·A > G·C/C·G. Binding isotherms for each complex fit well to an n = 1 Hill equation, supporting a 1:1 ligand:DNA stoichiometry (Fig. 3). To establish that the specificity was not position dependent, controls were performed on polyamide **1** at different Im (pAU16, 5′-AAAGAGAA**X**AG-3′) and β positions (pAU13, 5′-AAAGAG**X**AGAG-3′). The observed complex affinities and sequence selectivities were similar to that described above for pAU8 and pAU15, respectively (data not shown).

Polyamide Orientation at 5′-AAAXAXAAXAXAAA-3′ Sequences. Because Im binds all four base pairs at a single position, it would be interesting to ask what happens if one varies simultaneously the base pairs at all four Im positions. Quantitative DNase I footprint titration with **1** at 5′-AAA**X**A**X**AA**X**A**X**AAA-3′ (**X** = G, C, A, T) reveals a modest 3- to 5-fold preference for the **X** = G site (Fig. 4). Affinity cleavage analysis with polyamide **1** confirms a single orientation when **X** = G, consistent with 1:1 binding, which reverses when **X** = C, similar to observations made by Laemmli (13). Although the all A site (**X** = A) lacks a DNase I footprint because of a characteristic lack of cleavage at A tracts by DNase, the cleavage pattern of **1** is visible and is oriented to the 5′ side of this binding site, similar to **X** = G. The cleavage pattern is broader than that observed at the other sites, likely because of an ensemble of slipped complexes. The **X** = T site reveals cleavage, however, on both sides of the binding site and to different extents. Because of the similarity in DNA sequence at this site when read 5′→3′ or 3′→5′, it is likely that polyamide **1** is binding 1:1 at this site but with a slight preference for one orientation.

Discussion

This quantitative study helps to elucidate the current state of the art for 1:1 polyamide:DNA complexes and creates a baseline for

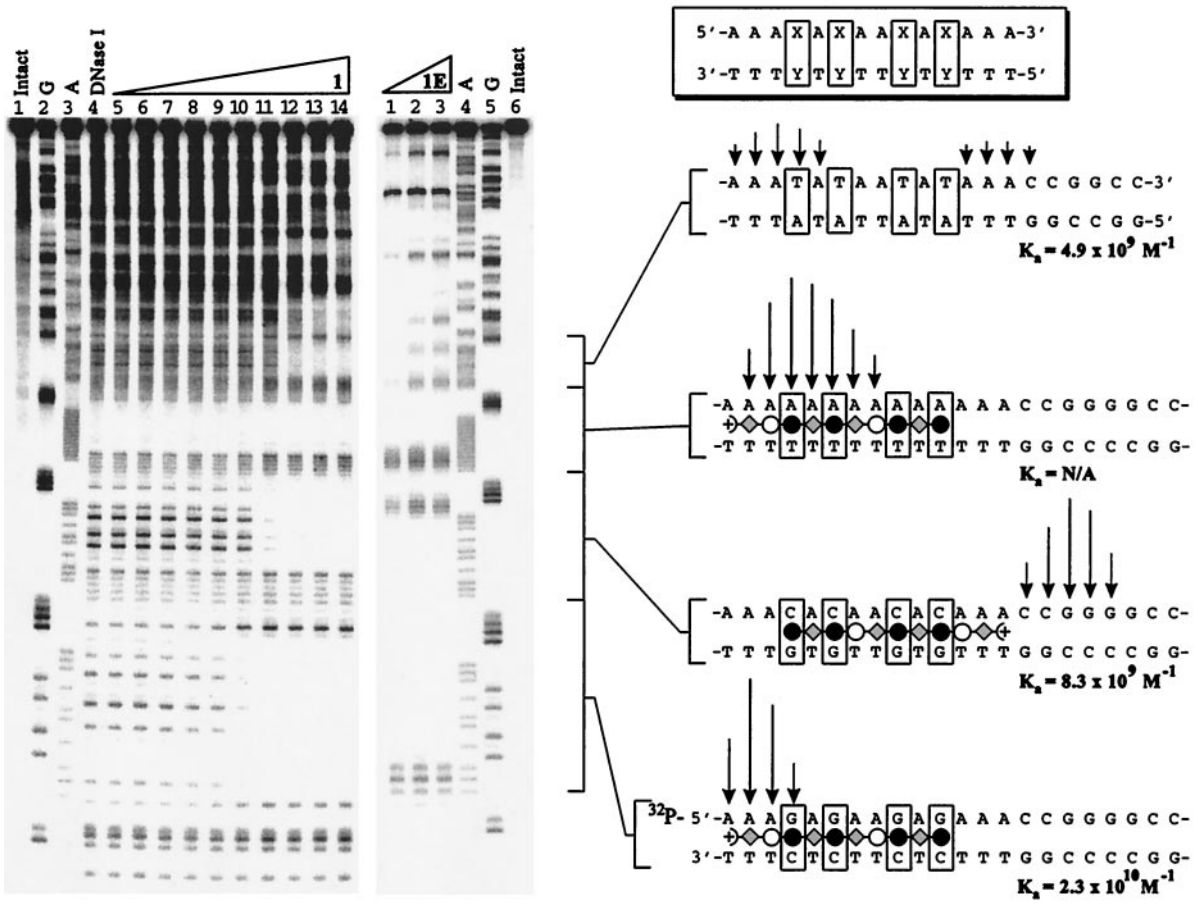

Fig. 4. (*Left*) Quantitative DNase I footprint titration experiment for polyamide **1** on the 298-bp, 5′-end-labeled PCR product from plasmid pAU18: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5–14, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM **1**, respectively. (*Center*) Affinity cleavage experiment with polyamide **1E** on the same PCR product of plasmid pAU18: lanes 1–3, 1 nM, 3 nM, and 10 nM **1E**, respectively; lane 4, A reaction; lane 5, G reaction; lane 6, intact DNA. (*Right*) Schematic illustrating the observed affinity cleavage patterns with arrows representing relative cleavage intensities. Polyamides are drawn as oriented, 1:1 complexes, as observed by affinity cleavage. Equilibrium association constants, K_a, are listed below each binding site.

future specificity studies (Table 2). An Im residue binds each of the four Watson·Crick base pairs with high affinity, whereas a β or Py residue prefers A,T over G,C base pairs but does not discriminate A from T. Steric inhibition between the exocyclic amino group of guanine and the Py and β residues may explain their A,T preference, as previously suggested by Dickerson and colleagues from x-ray structural analysis of 1:1 complexes (1). Based on the study of netropsin bound in a 1:1 complex with DNA, the promiscuous nature of the Im residue accepting G,C as well as A,T was anticipated (1). It is interesting to note that, although a single Im residue displays no significant base preference, the multiple G to C base mutation experiment demonstrates that the four Im residues taken together profoundly influence the orientation preference of the polyamide with respect to G·C vs. C·G base pairs (13). The similarity of orientation for **X** = G and A suggests that the purine-rich strand dominates the orientation preference and that polyamide C-N aligns in the minor groove 5′→3′ with the purine-rich strand. However, as observed for the mixed purine-pyrimidine tract (**X** = C), the polyamide orientation seems dominated by the G-rich strand.

Table 2. Recognition code for 1:1 polyamide:DNA complexes

X·Y	Im	β	Py	Hp
A·T	+	+	+	+
T·A	+	+	+	−
G·C	+	−	−	−
C·G	+	−	−	−

It is clear that the 1:1 and 2:1 binding stoichiometrics have quite different rules of recognition and specificity. This *ambiguity of sequence targeting depending on stoichiometry* raises a serious design issue for the field. For example, polyamide **1** binds four different sequences, 5′-AAA**X**A**X**AAA**X**A**X**AAA-3′ (**X** = G,C,A,T) with high affinity as a 1:1 complex and the sequence 5′-WGCGCWGCGCW-3′ (W = A or T) as a 2:1 complex according to the pairing rules (17). Therefore, a single polyamide binds multiple sites of similar size but quite different sequence composition, which is related to the stoichiometry of complexation in the minor groove. If specificity is the goal, this ambiguity raises the challenge to design next generation polyamides that enforce 2:1 vs. 1:1 binding in the minor groove (or the reverse). Recent examples are the "cycle motif" or "linked hairpins," which enforce the pairing rules (11, 22). Another direction will be the design of "integrated motifs," which combine polyamide modules for 1:1 and 2:1 binding modes. The "extended hairpin," comprising a six-ring hairpin module linked by a β-alanine to a three-ring pyrrole module, is an example of this class (23). One can imagine that integrated motifs could exploit the preference of 1:1 vs. 2:1 binding relevant to the sequence-dependent mi-

crostructure of DNA, i.e., certain sequence contexts with a normal β-like minor groove that may prefer pairs of rings (2:1 binding), whereas narrow minor groove tracts might allow the steric fit of only single rings (1:1 binding).

The unanticipated result of this study is the observation that an Hp residue in this polyamide sequence context can distinguish one of the four Watson·Crick base pairs, an A,T. Whether the hydroxyl moiety lies asymmetrically in the cleft between A and T and makes a specific hydrogen bond to the O2 of T, as observed for Hp/Py recognition of T·A (24), awaits structural studies. However, the fact that a single aromatic carboxamide residue can select one of the four Watson·Crick base pairs within the 1:1 motif is an encouraging step toward a set of rules for DNA recognition similar to the 2:1 motif. Whether three more aromatic residues can be invented to complete a 1:1 recognition code remains to be determined.

We thank the National Institutes of Health (Grant 27687) for research support and a training fellowship (to A.R.U.).

1. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1376–1380.
2. Pelton, J. G. & Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5723–5727.
3. Dervan, P. B. & Bürli, R. W. (1999) *Curr. Opin. Chem. Biol.* **3,** 688–693.
4. Youngquist, R. S. & Dervan, P. B. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2565–2569.
5. Kelly, J. J., Baird, E. E. & Dervan, P. B. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 6981.
6. Kielkopf, C. L., Baird, E. E., Dervan, P. B. & Rees, D. C. (1997) *Nat. Struct. Biol.* **5,** 104–109.
7. Youngquist, R. S. & Dervan, P. B. (1987) *J. Am. Chem. Soc.* **109,** 7564–7566.
8. Trauger, J. W., Baird, E. E., Mrksich, M. & Dervan, P. B. (1996) *J. Am. Chem. Soc.* **118,** 6160–6166.
9. Swalley, S. E., Baird, E. E. & Dervan, P. B. (1997) *Chem. Eur. J.* **3,** 1600–1607.
10. Mrksich, M., Parks, M. E. & Dervan, P. B. (1994) *J. Am. Chem. Soc.* **116,** 7983–7988.
11. Herman, D. M., Turner, J. M., Baird, E. E. & Dervan, P. B. (1999) *J. Am. Chem. Soc.* **121,** 1121–1129.
12. Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B. & Dabrowiak, J. C. (1986) *Biochemistry* **25,** 7408–7416.
13. Janssen, S., Durussel, T. & Laemmli, J. K. (2000) *Mol. Cell* **6,** 999–1011.
14. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboartory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
15. Baird, E. E. & Dervan, P. B. (1996) *J. Am. Chem. Soc.* **118,** 6141–6146.
16. Urbach, A. R., Szewczyk, J. W., White, S., Turner, J. M., Baird, E. E. & Dervan, P. B. (1999) *J. Am. Chem. Soc.* **121,** 11621–11629.
17. Dervan, P. B. & Urbach, A. R. (2000) in *Essays in Contemporary Chemistry*, From Molecular Structure Toward Biology, eds. Quinkert, G. & Kisakürek, M. V. (Verlag Helvetica Chimica Acta, Zurich).
18. Maxam, A. M. & Gilbert, W. S. (1980) *Methods Enzymol.* **65,** 499.
19. Iverson, B. L. & Dervan, P. B. (1987) *Nucleic Acids Res.* **15,** 7823–7830.
20. Taylor, J. S., Schultz, P. G. & Dervan, P. B. (1984) *Tetrahedron* **40,** 457–465.
21. Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986) *Methods Enzymol.* **130,** 132–181.
22. Herman, D. M., Baird, E. E. & Dervan, P. B. (1999) *Chem. Eur. J.* **5,** 975–983.
23. Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996) *Chem. Biol.* **3,** 369–377.
24. Kielkopf, C. L., White, S. Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B. & Rees, D. C. (1998) *Science* **282,** 111–115.